UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Ingegneria dell'Informazione
Curriculum: Controlli, Ottimizzazione e Sistemi Complessi (AOSC)

Integration of 3D imaging systems and augmented reality in the medical field

Candidate
Cosimo Aliani

Supervisors
Prof. Leonardo Bocchi

Prof. Carlo Colombo

Advisors
Prof. Fabio Schoen

Prof. Stefano Ricci

cycle XXXVIII, 2022-2025

Università degli Studi di Firenze, Dipartimento di Ingegneria dell'Informazione (DINFO).

A Emilio, Marco, Adriana, Leonardo ed Eva

Acknowledgments

I deeply acknowledge my supervisor, Prof. Leonardo Bocchi, who stimulated and helped me during the Ph.D. program. In addition, many thanks go to the colleagues of both the “EIDOLAB” and ”BMLAB” laboratories. All these acknowledged persons are with the University of Florence, Florence, Italy. Additionally, I want to thank all the people from the Imaginalis Srl company for their help and support.

Besides scholars at my home university, I warmly thank Prof. Klaudia Proniewska and Prof. Peter van Dam of the University of Kraków, Kraków, Poland, for their kind hospitality and support during my three-month stay in Kraków.

Finally, I am grateful to my grandfather Emilio Aliani, my parents Marco Aliani and Adriana Pitzalis, my brother Leonardo Aliani and my soul mate Eva Rossi for their trust in my skills and support.

Abstract

Background and rationale. Clinical imaging is increasingly expected to support spatial reasoning in context, yet most visualisation remains monitor-bound and operationally detached from the patient and the procedural environment. Augmented reality (AR) and optical 3D sensing (stereo depth cameras and LiDAR) are frequently presented as a single, monolithic technological stack; however, across real clinical workflows they are better understood as complementary modules whose coupling is optional and should be justified by added clinical value and by the robustness achievable on accessible hardware. This thesis was undertaken to reduce the translational gap between promising spatial-computing demonstrations and deployable tools by providing reproducible, open, and workflow-grounded building blocks.

Research question. The thesis asks: *How can AR and consumer-grade optical 3D imaging be engineered and evaluated, independently or in combination, to produce reliable, reproducible spatial-computing workflows for medical planning, training, and procedure-adjacent tasks, and what technical choices most strongly determine robustness in clinical-like conditions?*

Approach and contributions. I address this question through a programme of projects that collectively span sensing, calibration/registration, AI-derived anatomy, and user-facing AR interaction. First, I characterise the operational robustness of representative depth sensors (Intel RealSense D435 stereo and L515 LiDAR) under operating-room-like illumination and reflective surfaces, establishing practical constraints and selection criteria. Second, I benchmark geometric, feature-based, and marker-based registration methods for pointcloud alignment and adopt a ChArUco-guided strategy to achieve accurate, repeatable calibration and integration. Third, I develop end-to-end pipelines that bring precomputed volumes and automatic segmentations (TotalSegmentator and nnU-Net) into AR as manipulable 3D objects, enabling interactive interrogation of anatomy for planning, education, and rehearsal. Fourth, I demonstrate workflows in which depth sensing is tightly integrated: capturing external surface geometry and appearance to support patient positioning and to improve the realism of cranio-maxillofacial models by fusing external scans with computed-tomography-derived internal structures.

Key findings. Across the evaluated scenarios, stereo depth sensing (D435) proves more robust than LiDAR (L515) under adverse lighting and high-reflectance conditions, while ChArUco-guided alignment provides con-

sistent, repeatable registration compared to approaches relying solely on geometry or image features. The AR pipelines show that clinically meaningful interaction with volumetric imaging and AI segmentations can be achieved on portable, consumer-grade platforms, and that depth-sensor integration is most beneficial when external geometry or appearance is required (e.g., positioning, surface-aware fusion, or photorealistic texturisation), whereas AR-only visualisation suffices for many planning and educational tasks.

Conclusions and implications. Overall, the thesis argues for a pragmatic design paradigm for medical spatial computing: AR and optical 3D sensing should be treated as modular capabilities whose coupling is a decision as consequential as the underlying algorithms. The primary contribution is an open-source, portable, and reproducible blueprint, validated across multiple projects, that clarifies when and how 3D computer vision and AR enhance spatial reasoning, procedural training, and data-informed clinical decision-making. Remaining barriers include compute and latency budgets and incomplete workflow automation; clinical translation is best supported through edge/cloud co-processing, interaction refinement, and prospective studies aligned with regulatory and data-protection requirements.

Contents

Chapter 1

Introduction

Clinical decision-making is inherently spatial. Whether interpreting cross-sectional imaging, planning an intervention, positioning a patient, or navigating instruments in an operating room, clinicians must continuously translate complex three-dimensional (3D) anatomy into actionable steps. Yet the dominant interface for medical imaging remains largely two-dimensional: volumetric data are typically inspected on remote monitors through slices, projections, or static renderings. This mismatch, in which 3D tasks are mediated through 2D representations, can increase cognitive load and complicate hand-eye coordination, particularly in time-critical or precision-demanding procedures.

Two technological directions are reshaping this landscape. The first is *optical 3D imaging*, which captures real-world geometry and reconstructs spatial models of the patient or environment using depth sensing. The second is *augmented reality (AR)*, which enables clinicians to view and interact with digital content—such as volumes, segmentations, landmarks, and guidance cues—directly within the physical workspace. These directions are often presented as a single integrated pipeline; however, they do not constitute a mandatory stack. Depending on clinical intent, each can deliver value independently, and their coupling is beneficial only when integration yields measurable improvements in accuracy, usability, safety, or efficiency.

Optical depth sensing can support clinically relevant tasks such as non-contact surface measurement, quantitative morphology, patient positioning, and spatial referencing within imaging suites. In practice, systems span stereo vision, structured light, time-of-flight (ToF), and LiDAR, each with

distinct trade-offs in accuracy, cost, ease of deployment, and robustness to lighting and motion. More recent learning-based approaches broaden the design space by aiming to recover depth or scene structure from limited sensing, potentially lowering hardware barriers. However, clinical usefulness depends not only on reconstruction fidelity in ideal conditions, but also on calibration stability, repeatability across operators, resilience to real-world environments, and the ability to validate performance with clinically meaningful metrics.

AR promises a complementary capability: it can relocate imaging-derived information from a detached display into the point-of-care context, potentially improving spatial comprehension and supporting hands-free interaction. At the same time, deploying AR in clinical environments introduces hard requirements: robust tracking, reliable registration, ergonomic and hygienic constraints, compatibility with established workflows, and careful attention to safety and regulatory expectations. As a result, the success of AR is rarely determined by rendering quality alone; it depends on the full system pipeline, including interaction design and operational reliability.

Research gap and premise. A recurring challenge in the field is not merely how to build increasingly sophisticated 3D sensing or AR prototypes, but how to decide, based on clinical goals and constraints, which technology to use, in what configuration, and why. When does optical 3D imaging alone provide sufficient benefit? When does AR alone meaningfully improve the interpretation and communication of imaging data? And when does coupling the two justify the added complexity of integration, calibration, and validation? This thesis addresses these questions through a pragmatic stance: optical 3D imaging and AR are treated as distinct, self-standing lines of work, and they are deliberately coupled only when integration yields complementary capabilities and demonstrable clinical value.

Scope and aims. This doctoral work examines both the theory and practice of optical 3D imaging and AR for clinical purposes, with emphasis on low-cost, portable, and intelligent systems suited to real-world environments. Two complementary aims guide the work: to develop and evaluate optical 3D imaging and AR solutions as independent technologies, assessing the utility each can deliver in clinically motivated settings; and to demonstrate, where warranted, how their coupling yields complementary capabilities and greater

clinical value.

Contributions in brief. The thesis contributes methods and systems spanning acquisition, reconstruction, registration, and interactive visualisation. It includes experimental evaluations of depth-sensing and reconstruction strategies under practical constraints; pointcloud processing and registration approaches for spatial consistency; and AR applications for medical imaging interaction, training, and simulation. These contributions are instantiated in project-based case studies such as a non-ionising 3D scouting system for patient positioning in cone-beam CT (CBCT), a computed tomography (CT) slice simulation tool using ChArUco-based spatial referencing, and pipelines combining anatomical segmentation with AR-based visualisation.

Thesis organisation. Chapters 2 and 3 provide the scientific and technological background, reviewing optical 3D imaging and AR and positioning this work within the broader state of the art. Chapters 4 and 5 then present the original project-based contributions and evaluations. Finally, Chapters 6 and 7 synthesise findings, discuss limitations and translational considerations, and outline directions for future work. The project-based logic and the relationships among the components are summarised schematically in Figure 1.1.

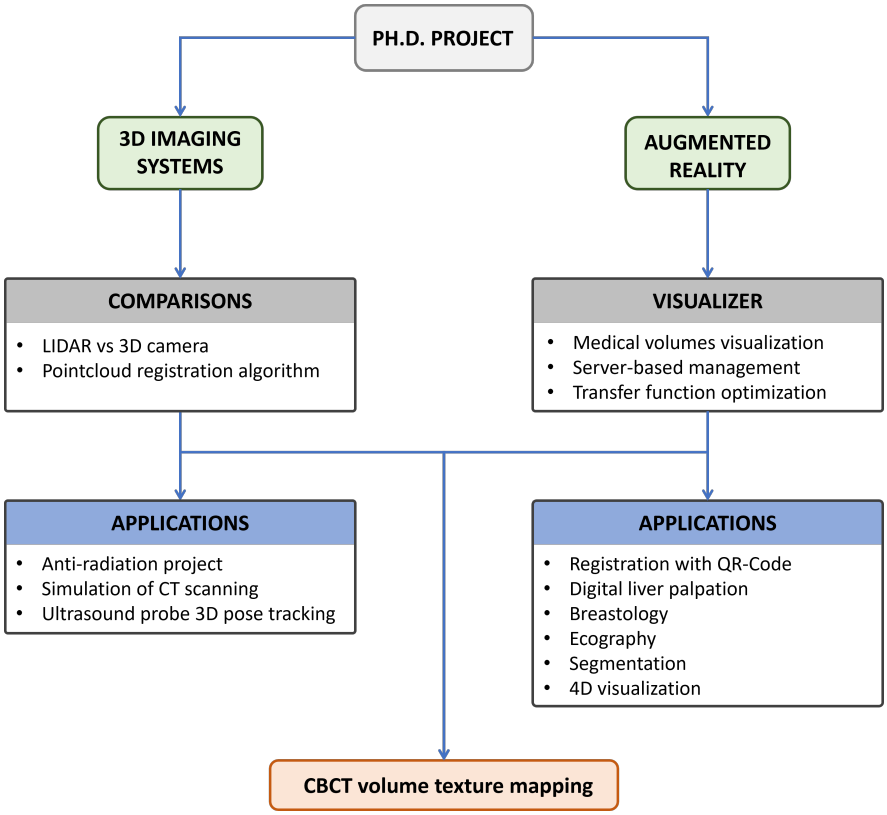


Figure 1.1: Schematic overview of the thesis structure.

Part I

State of the art

Chapter 2

Optical 3D imaging systems

This chapter surveys optical three-dimensional (3D) imaging systems, focusing on their historical development, physical principles, and classification. It also reviews recent advances driven by artificial intelligence and sensor fusion. Specific attention is given to clinical applications in diagnostics, surgery, and telemedicine, together with a critical analysis of current limitations.

2.1 Introduction

Optical three-dimensional (3D) imaging systems have transformed a wide array of scientific and clinical disciplines by enabling the reconstruction of external anatomy and scene geometry. In contrast to tomographic modalities (e.g., CT, MRI, ultrasound) that recover *internal* structures, optical 3D imaging reconstructs *external* surfaces and spatial layouts from light measurements. Compared with traditional two-dimensional (2D) views, optical 3D data provide richer spatial information, thereby supporting diagnostic precision, surgical planning, and therapeutic interventions. These systems exploit well-established optical principles—stereopsis, triangulation via structured illumination, and time-of-flight (ToF), including scanning LiDAR. Recent advances in sensor miniaturisation, computing power, and artificial intelligence further extend their capabilities, particularly through integration with augmented reality (AR) and real-time navigation tools. This chapter reviews the historical evolution, foundational technologies, and device typologies of *optical* 3D imaging systems, with a specific focus on biomedical

applications.

2.2 Historical background

The roots of optical 3D imaging trace back to nineteenth-century stereoscopic photography, where paired images conveyed depth via binocular disparity. On the optical side, rapid progress in the late twentieth and early twenty-first centuries brought high-precision triangulation sensors, structured-light scanners, and ToF cameras to practice, followed by consumer RGB-D devices and, more recently, compact LiDAR integrated into mobile platforms. While CT and MRI pioneered volumetric imaging of internal anatomy, the present chapter is restricted to *optical* systems that recover external surfaces; these play a complementary role in surgical navigation, pre-operative planning, prosthesis design, and telemedicine.

2.3 Types of optical 3D imaging systems

Optical 3D imaging systems can be broadly categorised by operating principle and sensor technology. The most prevalent types include:

- **Stereo vision systems:** two (or more) cameras at different viewpoints mimic binocular vision. Disparity between images yields depth via triangulation.
- **Structured-light systems:** a projector emits a known pattern (e.g., grids, stripes). A camera captures pattern deformation, which is decoded to recover depth.
- **Time-of-Flight (ToF) cameras:** the device emits modulated light and measures the round-trip phase or delay to infer distance in real time.
- **LiDAR (Light Detection and Ranging):** a scanning, laser-based ToF method renowned for range and accuracy; increasingly used for external body capture and in-room localisation.
- **Photogrammetry:** multiple overlapping photographs from different viewpoints are processed (Structure from Motion (SfM) or Multi-View Stereo (MVS)) to reconstruct 3D structure.

2.3.1 Emerging techniques in 3D imaging

Recent advances in computer vision, AI, and sensor fusion are extending optical 3D imaging beyond traditional hardware limits. Deep networks can infer disparity or depth from monocular or stereo imagery using supervised and self-supervised learning, improving robustness in low-texture or repetitive regions. Neural Radiance Fields (NeRFs) and related representations model scenes as continuous functions learnt from calibrated photographs, enabling photorealistic novel-view synthesis and implicit geometry; although still computationally demanding for intraoperative use, they are promising for pre-operative modelling and high-fidelity texturisation. Hybrid systems that fuse LiDAR, RGB(-D), and inertial measurement units (IMUs) improve robustness, resolution, and situational awareness—particularly valuable in dynamic clinical environments requiring real-time feedback.

2.4 Applications in medicine

Optical 3D sensing technologies, like RGB-D cameras, stereo systems, ToF sensors, and LiDAR, are increasingly adopted where real-time spatial awareness, surface reconstruction, and interactivity are valuable. Their non-ionising nature, portability, and low cost make them attractive for bedside diagnostics, surgical assistance, and training. Representative applications include:

- **Infant cranial deformation monitoring via smartphone photogrammetry:** a low-cost, smartphone-based photogrammetric approach was validated for assessing cranial deformation. Using structure from motion on slow-motion video acquired during routine paediatric consultations, surface models—when texture-enhancing fitted caps were used, achieving sub- to low-millimetre agreement with CT and MRI references, offering a radiation-free, accessible alternative for monitoring positional plagiocephaly [1].
- **Surface 3D scanning for personalised prosthetics and surgical planning:** photogrammetry, structured light, and stereo vision enable fast, non-contact capture of detailed surface geometries for custom implants, prosthetics, surgical guides, and epidemiological studies; their affordability and speed suit routine practice when internal imaging is unnecessary [2].

- **Automated body-weight estimation for emergency drug dosing:** RGB-D/ToF systems (e.g., Microsoft Kinect) have been explored as non-contact, AI-driven tools for estimating total body weight when conventional weighing is infeasible, often surpassing clinician estimates and anthropometric formulae; pointcloud features and digital twin models enable rapid, reliable inference [3].
- **LiDAR-enhanced 3D documentation in forensic autopsy practice:** mobile-device LiDAR enables in-room, high-fidelity 3D capture of cadaveric surfaces for accurate metric documentation, retrospective analysis, and court-admissible 3D prints—improving reproducibility over conventional 2D photography [4].
- **Optimising low-cost 3D body scanning with depth cameras:** consumer ToF and stereo sensors (e.g., Kinect v2, RealSense D435) mounted on rotating rigs with KinectFusion can deliver repeatable, sub-centimetric anthropometry at optimal ranges, suggesting viable, low-cost alternatives to commercial scanners [5].
- **Home-based AR rehabilitation using RGB-D motion tracking:** integrating an RGB-D camera with an interactive AR interface enables real-time limb tracking, joint-angle estimation, and posture correction, with telemonitoring for remote supervision—improving engagement and continuity of care [6].
- **RGB-D reconstruction and semantic mapping for assistive/-clinical technologies:** real-time scene mapping and semantic understanding support navigation aids, therapy supervision, and medical robotics; RGB-D sensors combined with SLAM and deep learning yield robust models in dynamic healthcare settings [7].
- **3D wound assessment via real-time RGB-D reconstruction:** commodity RGB-D cameras enable non-contact detection, segmentation, and 3D measurement of chronic wounds with sub-centimetric repeatability, integrating well with telemedicine workflows [8].
- **Facial analysis and biometrics through RGB-D depth sensing:** structured light, (active) stereo, and ToF improve robustness to lighting and pose for clinical/biometric facial tasks; a quality function deployment (QDF) based framework aligns sensor specifications

to application demands, with active stereo emerging as versatile across ranges [9].

- **Optical 3D imaging for soft-tissue modelling and forensic lesion analysis:** laser triangulation, structured light, and photogrammetry provide non-contact, high-resolution surface models for prosthetics, orthopaedics, dermatology, and medico-legal documentation, enabling quantitative, shareable evidence [10].
- **Sub-millimetre respiratory motion tracking for radiotherapy using RGB-D and PCA:** a real-time, non-contact system (e.g., Asus Xtion PRO) reconstructs respiratory motion in thoracic/abdominal regions with sub-millimetre precision, supporting gating and planning without radiographic imaging or wearable markers [11].
- **4D body modelling and VR feedback to enhance obesity-treatment adherence:** a low-cost network of RGB-D cameras (Intel RealSense) produces textured meshes and accurate anthropometrics; an immersive VR module visualises temporal change to increase engagement and adherence [12].

In summary, optical 3D sensing is proving to be a versatile, cost-effective tool across many areas of clinical practice. Its synergy with AR and AI facilitates real-time guidance and simulation and brings advanced spatial capabilities closer to the point of care.

2.5 Critical considerations and current limitations

Despite their growing relevance, optical 3D sensing systems—such as RGB-D cameras, LiDAR, stereo vision, and ToF sensors—face technical and practical constraints that affect depth accuracy, spatial resolution, clinical usability, and integration with downstream pipelines. Understanding these challenges is essential for designing deployable systems. Key limitations include:

- **Depth accuracy and spatial resolution:** many consumer grade sensors degrade on reflective, transparent, or low texture surfaces. Structured light and ToF can perform well under controlled lighting, but precision typically decreases with distance and can be affected by ambient IR.

- **Operating range and field of view:** structured light excels at short range but struggles beyond roughly one metre; LiDAR and passive stereo suit mid- to long-range capture but may lack fine detail at close distances. Narrow fields of view restrict usability in cluttered environments such as operating rooms.
- **Computational load and data quality:** real-time acquisition produces large data volumes. Filtering, registration, RGB–depth fusion, and learning-based post-processing can be computationally intensive, especially on embedded hardware.
- **Robustness in dynamic or cluttered scenes:** motion blur, occlusions, and variable illumination can destabilise depth estimates. Maintaining consistent measurements during patient and operator motion is challenging.
- **Device calibration and stability:** accurate reconstruction depends on reliable intrinsic/extrinsic calibration. Physical handling and temperature drift can misalign sensors, necessitating periodic recalibration that may be impractical in non-technical settings.
- **Cost and hardware trade-offs:** while affordability has improved, trade-offs remain among cost, resolution, robustness, and integration effort. High-end LiDAR offers excellent accuracy but may be expensive or less ergonomic than compact RGB-D cameras.
- **Lack of standardisation and interoperability:** heterogeneous formats of data and processing APIs hinder interoperability across devices and vendors, complicating integration into clinical software and AR platforms.

These limitations underscore the importance of selecting optical 3D solutions aligned with specific technical and clinical requirements, balancing accuracy, robustness, latency, usability, and cost.

Chapter 3

Augmented reality

This chapter provides an overview of augmented reality (AR) systems with a focus on their historical evolution, device categories, and integration into clinical workflows. It also examines recent advances in AR hardware and software, particularly those driven by wearable technologies and artificial intelligence. Specific emphasis is placed on medical applications—ranging from surgical guidance to education and rehabilitation—followed by a critical assessment of current limitations.

3.1 Introduction

Augmented reality (AR) is a rapidly evolving technology that enhances the real-world environment by superimposing digital information—such as images, sounds, and 3D models—onto the user's perception of the physical world. Unlike virtual reality (VR), which replaces the real world with a fully simulated environment, AR enriches the user's surroundings without detaching them from it. In medicine, AR offers powerful possibilities for visualisation, diagnosis, training, and treatment, making clinical processes more intuitive, interactive, and informed.

3.2 Historical background

The conceptual foundations of AR can be traced to the 1960s, when Ivan Sutherland developed the first head-mounted display (HMD), the "Sword of Damocles" [13,14]. AR took more recognisable forms in the 1990s, driven by advances in computer vision and graphical overlays. In medicine, early uses were largely experimental, involving image-guided surgery and simulation. The 2010s marked a turning point with the advent of powerful mobile devices, depth sensors, and wearable AR systems, enabling a transition from research labs to clinical pilots and selected routine applications.

3.3 Types of augmented reality devices

AR systems in medicine can be broadly categorised by hardware platform and how digital content is presented:

- **Head-mounted displays (HMDs):** Wearable devices, such as Microsoft HoloLens 2 and Magic Leap, project digital content directly into the user's field of view. They are widely used in surgery and medical training.
- **Handheld devices:** Smartphones and tablets with AR capabilities (e.g., Apple ARKit, Google ARCore) allow clinicians to visualise anatomical data or simulate procedures.
- **Projection-based systems:** Projectors overlay images directly onto physical surfaces, such as the patient's body or surgical field, providing intuitive guidance.
- **Mirror-based systems:** Also called spatial AR, these systems combine mirrors and projectors to align digital information with the user's view.
- **Contact-lens and retinal displays:** Still experimental, these technologies aim to deliver AR content via ultra-miniaturised displays on or within the eye.

3.3.1 Emerging techniques in AR systems

Recent advances have produced a new generation of AR technologies driven by machine learning, computer vision, and real-time sensing. AI-powered

object recognition, simultaneous localisation and mapping (SLAM), and environment-aware overlays allow AR systems to adapt dynamically to complex medical settings. Intraoperative AR is increasingly integrated with real-time imaging (e.g., MRI, ultrasound), enabling context-aware visualisation and dynamic registration of surgical instruments. Cloud-backed architectures facilitate remote collaboration, teleguidance, and large-scale training simulations. Haptic feedback and spatial audio are being incorporated to create multimodal AR experiences that enhance situational awareness.

3.4 Applications in medicine

AR technologies are increasingly integrated into clinical workflows, providing enhanced visualisation, spatial contextualisation, and user interaction across a wide range of applications. Representative areas include:

- **Performance and usability of tracked AR overlays in laparoscopic liver surgery:** A controlled usability study investigated an optically tracked AR system that overlaid pre-operative 3D models of liver anatomy and tumour targets onto laparoscopic video. Compared with baseline (no guidance) and standard 3D navigation, the AR overlay significantly improved localisation accuracy (e.g., median error reduction from 25.8 mm to 9.2 mm in surgeons), confidence, and perceived usability, with the greatest benefits for less experienced users. Minor limitations included video lag and occasional misalignment [15].

- **Tablet-based markerless AR for procedural training in emergency medicine:** A randomised controlled trial (Acidi et al.) evaluated a markerless, tablet-based AR system for chest-tube insertion on a mannequin torso. Participants trained with AR guidance achieved higher anatomical accuracy, fewer critical errors, and shorter completion times than controls. The system's portability and lack of fiducials suit low-resource or mobile training [16].

- **AR-enhanced navigation in neurosurgery:** AR systems superimpose segmented CT/MR data onto patient anatomy to improve intraoperative navigation. Core components include registration, real-time tracking, rendering, display hardware, and robotics. Despite issues such as brain shift, AR provides real-time anatomical context that can enable more precise resections [17].

- **Soft-tissue deformation tracking for AR-guided liver surgery:** By updating virtual models to match real-time organ deformation, AR enhances control over margins and ablation zones, improving needle placement accuracy in shifting surgical fields [17].
- **AR in orthopaedic implant alignment and joint kinematics:** AR overlays implant guides and kinematic vectors onto the surgical field, enabling patient-specific planning and real-time adjustment. Prototype systems combining pre-operative models with intraoperative AR feedback reduce positioning errors [17].
- **Vein visualisation for intravenous access using AR projection:** Systems such as AccuVein use near-infrared illumination with projection to enhance vascular visualisation, improving cannulation success in difficult cases (paediatric, geriatric, obese) [17].
- **Augmented reality in mental health and neurorehabilitation:** Multimodal AR systems (e.g., MindMaze) support cognitive rehabilitation by combining motion tracking, EEG input, and gamified feedback; AR "virtual contact" systems mitigate psychological stress in hospitalised children [17].
- **AR-based telementoring in surgery via wearable displays:** By streaming the surgeon's viewpoint to remote experts who annotate and guide in real time, AR expands access to expertise in resource-constrained settings and supports skill development [17].
- **Scoping review of AR in surgery: domains, displays, registration accuracy:** A PRISMA-guided review (Barcali et al.) of 34 studies (2019–2022) reported strong adoption in orthopaedics, maxillofacial surgery, and oncology; HoloLens (1/2) was the most common HMD, with marker-based rigid registration dominant. Registration errors ranged from sub-millimetre to a few millimetres; limitations included vergence–accommodation conflict and restricted field of view [18].
- **AR to enhance knowledge, procedural skills, and social learning in medical education:** A narrative review (Dhar et al.) reported that systems such as HoloHuman, OculAR SIM, and HoloPatient improved spatial understanding and procedural competence with fewer adverse effects than VR, and supported teamwork and communication in high-stakes simulations [19].

- **AR-guided pedicle screw navigation in spine surgery:** A cadaveric study (Felix et al.) using VisAR on HoloLens 2 achieved 96.0% accuracy (Gertzbein–Robbins grades A–B) for 124 screws, with mean angular and distance errors of 2.4° and 1.9 mm, while reducing fluoroscopy reliance [20].
- **AR for robotic liver surgery: from planning to intraoperative navigation:** A review (Giannone et al.) highlighted AR overlays for lesion localisation, vascular identification, and margin guidance, with hybrid approaches (AR + ultrasound/ICG fluorescence) to address soft-tissue registration [21].
- **Systematic evidence of AR impact in medical education:** An umbrella review (Tene et al.) of 28 empirical studies found consistent improvements in procedural accuracy, decision-making, and learner engagement with AR, though more standardised validation is needed [22].
- **Open-source AR-guided surgery using OST-HMDs in cadaveric maxillofacial procedures:** An open-source HoloLens system (Puladi et al.) for zygomatic arch fracture reduction demonstrated comparable accuracy to conventional methods and high usability, offering a replicable pathway to clinical translation [23].
- **AI-assisted real-time de-occlusion for AR in robotic surgery:** A first-in-human study integrated deep segmentation with 3D model overlays (Holoscan SDK) to manage occlusions across three live cases, improving depth perception and reducing perceptual latency [24].
- **Mixed-reality holographic visualisation for congenital heart surgery:** A case series (Morimoto et al.) used HoloLens 2 with remote rendering to display patient-specific models, improving spatial understanding and team communication in complex congenital cases [25].
- **Global research trends and clinical hotspots in AR:** A bibliometric analysis (Yeung et al.) mapped over 8,000 publications, with major clusters in surgery, neurorehabilitation, pain management, and education, and highlighted challenges in standardisation and long-term benefit [26].
- **Head-mounted AR for ultrasound-guided central line placement:** A phantom study (Sun et al.) combining ultrasound with

HoloLens improved targeting accuracy, reduced head/eye shifts, and enhanced efficiency while maintaining sterile technique [27].

- **Five-year landscape review of AR in medicine:** A systematic review (Eckert et al.) of 338 original studies (2012–2017) reported most applications at TRL 6–7, with limited clinical trials, underscoring the need for regulatory validation and standardised metrics [28].
- **Extended reality in mitral valve surgery:** A systematic review (Nanchahal et al.) found that XR (AR/VR) supports pre-operative planning, ring selection, suture strategies, and outcome prediction through interactive, patient-specific models [29].
- **Validation status and scope of AR applications in training:** A review (Barsom et al.) identified validated AR platforms for laparoscopic and neurosurgical training but noted the absence of full predictive validation linking simulator performance to clinical outcomes [30].
- **Augmented reality for situated, whole-task learning:** An early conceptual paper (Kamphuis et al.) argued that AR affords situated, multimodal training with real-time feedback and reduced cognitive load, and outlined a research agenda for rigorous evaluation [31].

At the same time—and given this thesis's focus on AR alongside optical 3D imaging—there are domains where their synergy is explicit. Selected examples include:

- **Projection-based AR for neurosurgical modelling:** Chien et al. [32] projected CT-derived vascular models onto curved head phantoms, correcting for surface curvature and observer viewpoint (tracked by RGB-D) to achieve < 2.1 mm positional deviation via improved ICP and face tracking—supporting simulation and guidance without wearables.
- **Mobile AR rehabilitation with RGB and LiDAR:** Kaewrat et al. [33] combined MediaPipe (RGB) with ARFoundation (LiDAR) for posture tracking during "marching-in-place" exercises. LiDAR offered higher tracking accuracy in complex environments, while RGB improved usability with front-facing displays.
- **Calibrating RGB-D with a mobile C-arm:** Wang et al. [34] introduced a phantom visible to both modalities to compute a precise

3D–2D projection matrix, enabling markerless registration to intraoperative X-rays and AR overlays from the C-arm perspective (RMSE ≈ 0.54 mm).

- **Markerless 3D–3D calibration between RGB-D and CBCT:** Lee et al. [35] mounted a depth camera on a mobile C-arm and registered surface pointclouds to CBCT volumes using FPFH and ICP, achieving mean TRE of 2.58 mm and robust visualisation from arbitrary viewpoints.
- **On-patient AR visualisation of pre-operative images:** Wu et al. [36] registered RGB-D facial capture to CT using an Improved ICP with stochastic perturbations and weighting, enabling HoloLens overlays with sub-3 mm error on a head phantom.

These examples illustrate how integrating optical 3D sensing (e.g., RGB-D, LiDAR) with AR can enhance planning, navigation, telerehabilitation, and education.

3.5 Critical considerations and current limitations

Despite growing adoption, medical AR faces technical and practical challenges that limit generalisability and robustness:

- **Tracking accuracy and calibration:** Precise alignment of overlays with the physical world remains difficult, especially in dynamic or deformable scenes (e.g., soft tissue).
- **Field of view and resolution:** Many AR devices have limited display areas and modest angular resolution, which constrains usability for complex tasks.
- **Ergonomics and user fatigue:** Extended use of HMDs may cause discomfort or fatigue, particularly during long procedures.
- **Data integration and registration:** Real-time fusion of preoperative images (CT/MRI) with the surgical field requires robust, maintainable registration; intraoperative drift and deformation are persistent issues.

- **Cost, scalability, and interoperability:** Devices remain relatively expensive; heterogeneous platforms and APIs hinder standardisation and large-scale deployment.
- **Privacy and cybersecurity:** Cloud connectivity and use of patient data demand stringent safeguards against unauthorised access and breaches.

Careful evaluation of these factors is essential to ensure AR systems are deployed effectively, ethically, and in alignment with clinical constraints.

Part II

Projects developed during this thesis

Chapter 4

Optical 3D imaging systems projects

Chapter scope, rationale, and structure

This chapter reports the optical 3D-imaging projects developed during the doctoral programme, with an emphasis on *system-level choices* (sensor modality and integration constraints), *geometric processing* (registration and pose estimation), and *clinical translation drivers* (workflow impact, accuracy targets, and failure modes). Unlike a literature review, the chapter is organised around original experimental pipelines, quantitative evaluations, and prototype integrations that were implemented and tested on real devices and imaging systems. Where relevant, prior work is cited to contextualise design choices; however, the primary focus is the methodological and experimental contribution of the presented projects.

The overarching rationale is pragmatic: in medical environments, optical depth sensing is attractive because it is non-ionising, relatively low-cost, and potentially deployable at the point of care. Its utility, however, depends on multiple factors, for example, robustness to challenging materials and lighting and reliable registration into clinically meaningful frames.

The chapter is organised as a roadmap of projects carried out during the PhD, designed to present a progressive, unified line of research. The first section addresses the foundational choice of the sensing technology by comparing stereo and LiDAR depth cameras in conditions relevant to medical settings. The second section evaluates alternative point-cloud registration

and pose-estimation strategies, establishing the methodological basis for reliable 3D integration. Subsequent sections then present application-oriented projects in which the selected device and the validated registration pipeline are embedded into specific biomedical workflows, illustrating an incremental transition from technological evaluation to clinically motivated prototypes.

4.1 3D camera and LiDAR: a comparative evaluation for medical applications

In the early stages of this doctoral research, we conducted a comparative study to evaluate two depth-sensing technologies—the Intel RealSense D435 (stereo depth camera [37]) and the Intel RealSense L515 (LiDAR-based sensor [38])—for prospective use in medical imaging and analysis. Both devices are shown in Figure 4.1. The objective was to assess depth accuracy, material sensitivity, and robustness under environmental and surface conditions representative of clinical or preclinical scenarios. A concise technical overview of both devices is reported below:

- **Intel RealSense D435**
  - *Technology*: active infrared stereo vision
  - *Working range*: approximately 0.2–10 m (scene dependent)
  - *Depth resolution*: up to 1280×720 at 30 fps (higher frame rates at reduced resolution)
- **Intel RealSense L515**
  - *Technology*: solid-state LiDAR (direct time-of-flight)
  - *Working range*: approximately 0.25–9 m (scene dependent)
  - *Depth accuracy*: high precision at short range (millimetre-level in favourable conditions)

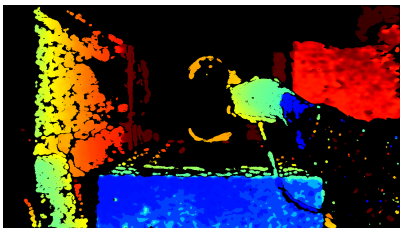
We evaluated a selection of materials spanning diverse optical and mechanical properties:

- **Soft plastic catheter (cannula)**: flexible, medical-grade tubing used for fluid administration; its translucency and small, curved geometry challenge depth recovery.

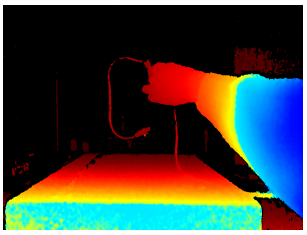
(a) (b)

Figure 4.1: The two depth-sensing devices evaluated during the preliminary study. (a): Intel RealSense D435; (b): Intel RealSense L515.

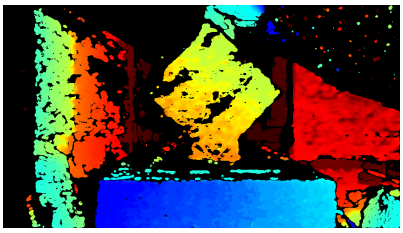
- **Reference tool with retroreflective markers**: a rigid frame with spherical retroreflectors used for optical tracking; included to test capture of discrete targets under IR illumination and to assess visibility/-contrast.
- **Stainless steel (matte/grounded)**: a common, highly reflective surgical material; used to probe specular reflections and glare.
- **Matte carbon**: low-reflectance, fibrous surface; stresses stereo correspondence on fine texture.
- **Glossy carbon**: similar to matte carbon but reflective; combines texture with specular highlights.
- **Fabrics (clothing)**: soft, deformable, low-contrast surfaces; proxy for drapes/skin-like behaviour.
- **Transparent plastic**: transmissive material; probes IR transmission and scattering effects.
- **Transparent plexiglass**: rigid, optically clear; used to investigate refraction, internal reflections, and depth voids.
- **White plastic**: diffuse, uniform baseline for estimating nominal depth noise and bias.



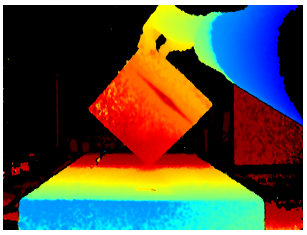
(a) Soft plastic catheter acquired with the Intel RealSense D435.



(b) Soft plastic catheter acquired with the Intel RealSense L515.



(c) Matte stainless-steel plate acquired with the Intel RealSense D435.



(d) Matte stainless-steel plate acquired with the Intel RealSense L515.

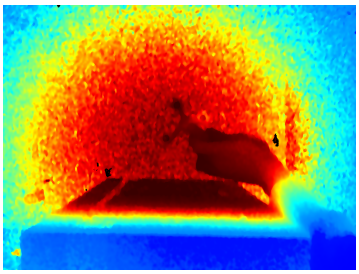
Figure 4.2: Depth images captured with both devices.

Representative depth images for selected objects are shown in Figure 4.2.

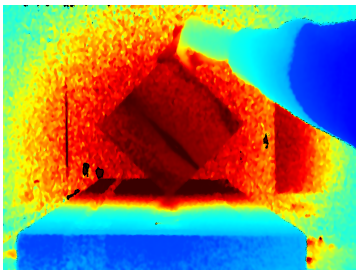
Two materials posed particular challenges for the LiDAR-based L515: stainless steel and the retroreflective tracking tool.

In both cases, intense backscatter drove the photodetector into saturation, producing a glare-like effect with extended regions of invalid depth (Figures 4.3(a)–4.3(b)). This behaviour is consistent with overexposure and blooming in direct time-of-flight systems confronted with specular or retroreflective returns.

These observations highlight a critical limitation of the L515 when exposed to highly reflective or retroreflective materials: saturation severely compromises depth integrity and coverage, which is problematic for scenarios involving metallic instruments or optical tracking markers. By contrast, the stereo-based D435 exhibited stable behaviour under the same conditions, with no visible saturation artefacts or large-scale data voids. On this basis, and given its broader robustness across materials, the D435 was selected as



(a) Glare/saturation with a surgical navigation tool bearing IR retroreflective markers.



(b) Glare/saturation with a matte stainless-steel plate.

Figure 4.3: Optical glare (sensor saturation) observed with the Intel RealSense L515 on reflective/retroreflective targets.

the primary sensor for subsequent stages of this research.

4.2 Pointclouds registration algorithms

Accurate alignment of multiple pointclouds into a unified reference frame is fundamental to 3D reconstruction and spatial mapping. In medicine, reliable registration enables pre-operative model alignment, intra-operative instrument tracking, and dynamic patient-surface mapping—thereby supporting navigation, monitoring, and reconstruction.

Robust, precise registration in clinical settings is challenging: depth data may be noisy or incomplete; surfaces can be reflective, transparent, or low-texture; viewpoints are constrained by ergonomics and sterility; and real-time requirements limit algorithmic complexity. Variability in illumination and appearance further complicates the pipeline.

To address these challenges, three distinct pointcloud registration strategies were developed and experimentally evaluated. Each was selected for its theoretical suitability to medical applications and implemented on real data captured with the Intel RealSense RGB–D camera:

1. a geometry-based technique that detects spherical fiducials via the Hough transform [39];

2. a feature-based pipeline that detects and matches keypoints using SIFT

descriptors [40];

3. a marker-based method employing a ChArUco board for high-precision pose estimation using hybrid visual features.

Each approach was tested under controlled yet realistic conditions to assess robustness, repeatability, sensitivity to noise and occlusion, and the accuracy of the recovered rigid-body transformations. The following subsections describe the principles, implementation details, and observed performance, with a critical analysis of suitability for clinical use.

4.2.1 Sphere-based registration algorithm using the Hough transform

The first strategy exploits geometric detection of spherical fiducials via the Hough transform [39], well-suited when the spatial configuration of identifiable primitives is known a priori. The experimental setup included passive, retroreflective spherical markers, common in optical surgical navigation, rigidly arranged in a calibrated pattern, visible in both RGB and depth channels.

The Hough transform maps image evidence into a parameter space in which peaks correspond to candidate shapes (lines, circles, spheres). In the 2D case used here, circle detection proceeds on the RGB image: edge pixels vote for circles parameterised by centre and radius; prominent peaks indicate likely circles. Detected circles are then back-projected to 3D by associating the corresponding pixels with the synchronously acquired depth map.

Because the physical layout (including pairwise inter-marker distances) was known, we solved for the rigid transformation (rotation and translation), aligning the detected 3D marker configuration to the reference model, and applied this to register the current pointcloud to the canonical frame.

The phantom comprised eight infrared-retroreflective spheres (diameter 10 mm) affixed to a lightweight frame. Crucially, markers were intentionally non-coplanar, improving the conditioning of the absolute orientation problem and increasing robustness to noise and partial occlusion. Figure 4.4 shows frontal and lateral views, highlighting the 3D distribution.

Limitations. Despite strong geometric constraints, several practical issues emerged. First, false positives: a circle Hough transform detects circular image structures regardless of semantics, frequently responding to background

(a) (b)

Figure 4.4: Frontal and side views of the spherical-marker phantom.

roundings or hardware features. Although radius priors help, they are specified in pixels and are not invariant to scale, leading to sensitivity to distance and perspective. Second, intermittent detection: partial occlusion, low illumination, or grazing angles yielded missed detections; the loss of even one sphere could destabilise or invalidate the rigid transform. Third, visibility: small, reflective markers were not consistently observable from arbitrary viewpoints. Together, these factors made the method insufficiently robust and repeatable for the intended context, and it was therefore excluded from subsequent pipelines.

4.2.2 Keypoint matching and feature-based registration algorithm

The second strategy estimates the relative pose between two RGB-D acquisitions by exploiting visual features that can be reliably re-identified across images. Intuitively, rather than aligning entire pointclouds directly, the method first searches for a set of distinctive image locations (called *keypoints*) that act as landmarks. These landmarks are then matched between two views and, because depth is available, each matched pixel can be converted into a 3D point. The final 3D registration is obtained by finding the rigid transformation that best aligns the resulting 3D correspondences. In more detail, a keypoint is a pixel location associated with a locally distinctive pattern, typi-

cally a corner or a textured region in which image intensity varies in multiple directions. Such locations are preferred because they can be detected repeatedly even if the object is seen from a slightly different viewpoint. Once a keypoint is detected, it is described numerically by a *descriptor*: a compact vector summarising the appearance of the image patch around that point (e.g., local gradients). In this work, we used the Scale-Invariant Feature Transform (SIFT) [40], which was designed to make descriptors comparatively robust to changes in scale and rotation.

Given two images, SIFT keypoints and descriptors are computed independently for each view using the OpenCV library [41]. Each descriptor in the source image is then compared to descriptors in the target image to find the most similar candidates. However, raw nearest-neighbour matching often produces ambiguous correspondences, especially when the scene contains repeated textures or weak visual structure. To reduce false matches, we applied Lowe's ratio test: for each source descriptor, the best match and the second-best match are identified; the correspondence is accepted only if the best match is substantially better than the second-best. Formally, let d_1 and d_2 be the descriptor distances to the first and second nearest neighbours in the target image. A match is kept if:

$$\frac{d_1}{d_2} < L_{\text{thr}}, \tag{4.1}$$

with $L_{\text{thr}} = 0.7$ in our experiments. This criterion discards matches that are not clearly distinctive, which is crucial for avoiding gross registration failures.

For each surviving keypoint correspondence, the associated depth value is read from the synchronised depth map. Using the camera intrinsics, the 2D pixel coordinates (u, v) and depth z are back-projected into a 3D point $\mathbf{p} = (x, y, z)^\top$ in the camera coordinate system. Repeating this for both views yields a set of paired 3D points $\{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^{N}$.

The rigid transformation that aligns the source points to the target points is parameterised by a rotation $\mathbf{R} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$. It is estimated by minimising the least-squares discrepancy across all correspondences:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|^2, \qquad \text{s.t. } \mathbf{R}^\top\mathbf{R} = \mathbf{I},\ \det(\mathbf{R}) = 1. \tag{4.2}$$

This produces the transformation that best explains the matched 3D land-

marks in a global least-squares sense, and it can then be applied to register the full point cloud of one view onto the other.

Limitations. Although conceptually simple, the method proved unreliable in our target scenarios due to two fundamental issues. Keypoint detectors require local intensity variations (texture) to produce stable landmarks. Many clinically relevant surfaces—skin regions with weak texture, sterile drapes, matte plastics, and metallic instruments—are visually uniform or exhibit specular highlights, yielding too few repeatable keypoints. In these cases, the algorithm either fails to find enough correspondences or relies on unstable ones, which directly degrades the pose estimate. Additionally, even when keypoints are detected, their descriptors are not perfectly invariant: changes in distance, perspective, partial occlusion, and lighting can alter the local appearance, so that the same physical point is not consistently detected (or is mismatched) across views. This leads to sparse or inconsistent correspondences and, consequently, to unstable least-squares solutions with occasional large outliers. In a biomedical workflow, these failure modes translate into intermittent loss of registration and unpredictable misalignments. For this reason, despite satisfactory performance on well-textured test scenes, this feature-based approach was not adopted as a core component for downstream prototypes, where repeatability and robustness were prioritised.

4.2.3 ChArUco marker-based registration algorithm

The third strategy uses ChArUco markers [42], hybrid fiducials combining ArUco IDs with a chessboard, to achieve indexed detection with subpixel-accurate corners. This yields high-precision pose estimates suitable for calibration and registration. The combination of a chessboard and ArUco IDs, leading to a ChArUco board, is represented in Figure 4.5.

During this thesis, a rigid 7×7 plastic ChArUco board (shown in Figure 4.6), was used as the reference object. Each chessboard square measured 15 mm, and each ArUco marker measured 10 mm. The board was visible from both viewpoints of the RGB–D camera; detections were performed independently in the two RGB frames using OpenCV, with subpixel refinement of internal corners. Pose estimation routines then produced the relative rotation and translation between views. Applying the resulting rigid-body

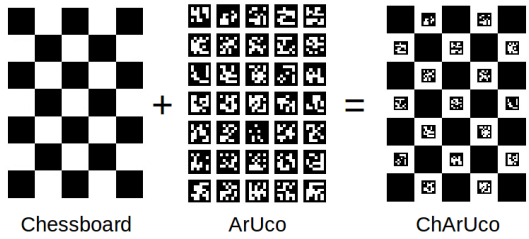


Figure 4.5: Example ChArUco pattern [43].

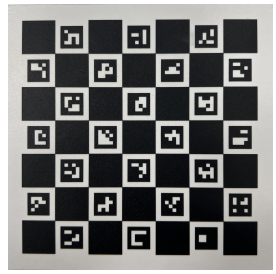


Figure 4.6: ChArUco board used in this project.

transform to one pointcloud aligned the two clouds, without reliance on natural scene texture.

Unlike the previous methods, ChArUco-based registration delivered consistently high detection stability and registration accuracy. Provided the board was at least partially visible in both views, the recovered transforms were accurate and repeatable. The method is largely insensitive to scene texture, tolerates moderate viewpoint changes, and integrates cleanly with RGB–D acquisition. For these reasons, it was selected as the reference approach for subsequent experiments.

However, when ChArUco is used beyond static inter-view registration—namely as a continuous pose-tracking cue to support motion-aware imaging or navigation—the requirements become more stringent. In these scenarios, the target must remain detectable over extended sequences and across changing viewpoints, and the pose stream must be not only accurate on average but also highly repeatable from frame to frame, because jitter can propagate into downstream compensation or guidance steps. Moreover, practical deployments impose non-ideal constraints on camera placement (available

mounting locations, safety clearances, occlusions, and permissible working distances), which makes pose quality strongly dependent on the camera–board geometry.

A representative example considered in this work is standing CBCT, where motion compensation can substantially benefit image quality in the presence of residual subject motion. In particular, standing CT/CBCT examinations performed on sedated horses provide an operating regime in which motion is reduced but not eliminated, and acquisition-time posture adjustments can still lead to view-to-view inconsistencies. For this reason, we investigated how viewing angle and working distance jointly affect ChArUco-based pose estimation, to identify camera placements that are robust under realistic acquisition constraints.

Optimising camera–ChArUco geometry for robust pose tracking in standing equine CBCT In marker-based optical tracking, pose quality is not only a property of the detection algorithm, but also of the camera–marker geometry. This consideration becomes particularly relevant in standing CT/CBCT acquisitions performed on sedated horses. In these examinations (shown in Figure 4.7), sedation is essential to ensure safety and tolerability; however, it does not eliminate motion. Residual postural sway, slow weight shifting, and intermittent head–neck adjustments remain common, and their amplitude is often small but persistent over the acquisition time. Because CBCT reconstruction integrates hundreds of projections acquired sequentially, even modest, view-to-view motion can accumulate into inconsistent projection geometry, manifesting as blurring and streak artefacts that degrade diagnostic interpretability. Within this scenario, an auxiliary RGB camera observing a rigidly attached ChArUco target provides a practical route to estimate frame-by-frame rigid motion for subsequent compensation. Importantly, the limiting factor is frequently not gross tracking failures, but pose jitter and geometric sensitivity: small estimation fluctuations can translate into unstable correction transforms and thus cap the achievable artefact reduction. For this reason, we investigated how viewing angle and working distance jointly affect ChArUco-based pose estimation, to identify camera placements that are robust in the specific operating conditions of sedated equine standing acquisitions.

The board was observed using the RGB module of the Intel RealSense D435, and pose estimation was performed with OpenCV routines for ChAr-

Figure 4.7: Illustrative standing equine CBCT workflow in which a lightweight ChArUco board is rigidly attached near the region of interest and observed by an auxiliary RGB camera for external pose tracking.

Uco detection and corner refinement. The relative transformation between camera and board was decomposed into *camera-to-plane angles* and *camera-to-plane positions*. Specifically, the angular parameters were $CtP_A = [\varphi, \zeta]$, where φ denotes the angle between the camera optical axis and the board normal (i.e., a viewing-angle parameter analogous to a yaw-like deviation from frontal view), and ζ represents the in-plane rotation about the board normal. Translational parameters were $CtP_P = [T_x, T_y, T_z]$, describing the position of the board origin expressed in the camera reference system. In the intended standing-acquisition scenario, the dominant components are expected to be φ and T_z, reflecting sway-driven changes in viewpoint and distance rather than large in-plane rotations. A visual representation of both CtP_A and CtP_P parameters is shown in Figure 4.8.

To isolate the effect of geometry, two dedicated fixtures were employed. First, to study the influence of viewing angle φ, the board was mounted on a motor-driven rotation wheel with 1° positioning resolution. The board origin was aligned with the wheel rotation axis to minimise parasitic translations and undesired rotation about ζ during angular sweeps. Second, to study the

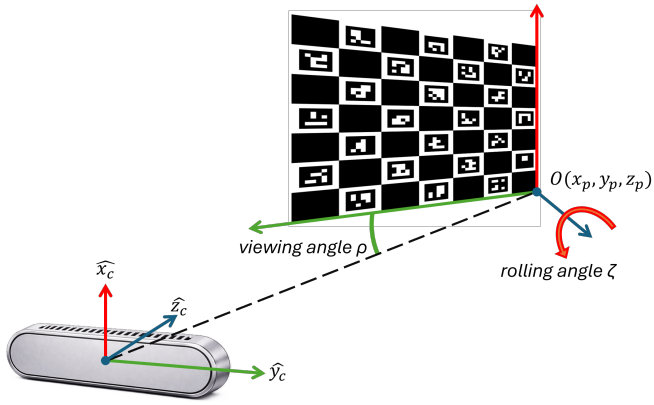
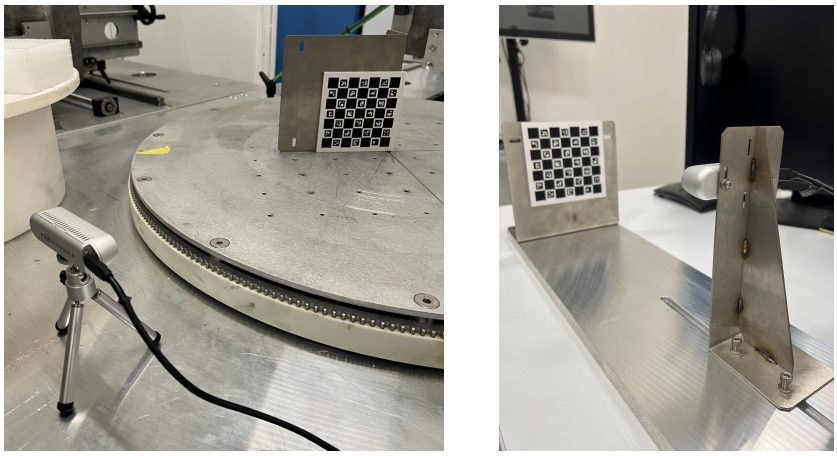


Figure 4.8: ChArUco board used in the experiments and definition of the camera/board reference systems. The pose is parameterised by camera-to-plane angles $CtP_A = [\varphi, \zeta]$ and camera-to-plane positions $CtP_P = [T_x, T_y, T_z]$. ChArUco's origin $O(x_p, y_p, z_p)$ is expressed in the camera reference system.

influence of working distance, the camera–board separation was adjusted along the optical axis using a 1 m linear rail system; the distance $T_{z,\mathrm{GT}}$ was measured with a ruler at each tested position. For angular experiments, the wheel-set angle provided φ_{GT}; to reduce backlash effects, target angles were approached consistently from the same direction.

Accuracy was quantified as absolute error with respect to ground truth, e.g. $e_\varphi = |\varphi - \varphi_{\mathrm{GT}}|$ and $e_{T_z} = |T_z - T_{z,\mathrm{GT}}|$. Precision (repeatability) was quantified as the standard deviation across repeated estimates acquired under identical conditions. For each run, a reference pose was first recorded and subsequently subtracted, so that the analysis focused on pose increments rather than absolute values, thereby reducing sensitivity to constant offsets.

A first experiment assessed whether repeated repositioning introduces additional variability beyond intrinsic camera/estimator noise—a relevant distinction for projection-wise tracking, where each projection corresponds to a single pose sample while geometry changes between views. With the camera fixed at $T_z = 40$ cm, poses were acquired at $\varphi \in \{45°, 25°, 5°, 0°, -5°, -25°, -45°\}$ using two modalities: (i) *continuous* acquisitions (ten estimates per angle without moving the wheel between repeats), and (ii) *cyclic* acquisi-



(a) (b)

Figure 4.9: Experimental setups for the pose-estimation study. (a): Viewing angle evaluation setup; (b): T_z evaluation setup.

tions (one estimate per angle, sweeping across angles and repeating the full cycle ten times). The two modalities yielded comparable errors, indicating that wheel repositioning did not substantially amplify variability under the adopted operating conditions.

The second experiment characterised how pose quality changes across viewing angles in a CBCT-motivated configuration. Angles from $+60°$ to $-60°$ were tested (including intermediate values), at two fixed working distances ($T_z = 40$ cm and $T_z = 60$ cm). For each (φ, T_z) condition, twenty repeated pose estimates were acquired. A consistent accuracy–precision trade-off emerged for rotation: the mean absolute error on φ was minimal near frontal view ($\varphi_{\mathrm{GT}} \approx 0°$) and increased progressively at more oblique views, whereas the standard deviation was highest near frontal view and decreased as obliquity increased. In other words, frontal configurations tended to minimise bias but maximise jitter, while moderately oblique configurations reduced jitter at the cost of increased bias. A similar trend was observed for the in-plane rotation ζ. Changing working distance also influenced this balance: increasing T_z tended to increase variability (particularly for angular parameters), consistent with the reduced marker footprint in the RGB image and the resulting sensitivity of corner localisation to pixel-level noise.

Finally, the camera–board distance was varied from $T_z = 25$ cm to $T_z = 60$ cm in 5 cm increments, acquiring twenty repeated estimates per position. Lateral components (T_x, T_y) remained stable across distances (low variability), whereas depth repeatability degraded with increasing distance: the standard deviation of T_z increased markedly as the board moved further away, even when the mean absolute error did not exhibit a strong monotonic trend. This indicates that, within the explored range, distance primarily impacts the repeatability of depth estimates rather than introducing a systematic bias.

Practical implications for experimental and clinical deployments. These findings suggest that camera placement should be selected according to the relative importance of bias versus jitter in the downstream application. In projection-wise motion compensation, frame-to-frame jitter can propagate into correction transforms and limit achievable artefact reduction; therefore, a strictly frontal configuration—although optimal in terms of mean angular error—may be suboptimal when repeatability is prioritised. Instead, a moderately oblique nominal viewpoint may provide more stable pose streams, provided that the induced bias remains compatible with the intended compensation model. In addition, excessive working distances should be avoided to preserve a sufficiently large marker footprint and maintain repeatable depth (T_z) estimation. Detailed numerical results for each test condition, together with additional methodological details, are reported in the preprint by Aliani et al. [44] (the peer-reviewed paper is under review).

4.3 Non-ionising optical scouting for CBCT positioning

In current clinical practice, cone-beam computed tomography (CBCT) systems often employ low-dose ionising “scout views” to verify the positioning of patients or objects on the acquisition table. This preliminary step ensures that the region of interest is aligned within the imaging volume before the full scan. Although each scout involves minimal dose, repetitive use—especially in sensitive populations or high-throughput environments—raises concerns about cumulative exposure and patient safety. Scout acquisitions also introduce workflow friction: because ionising radiation is emitted, the operator must exit the shielded room for each scout, increasing procedure time and

disrupting continuity. Reducing the frequency of these steps could improve both safety and operational efficiency in CBCT workflows.

To mitigate these issues, this project investigates a non-ionising, optical alternative for positional scouting by mounting a 3D depth camera directly on the CBCT gantry. Specifically, we evaluate the Intel RealSense D435 (stereo RGB–D) as a real-time spatial sensing module capable of indicating whether the object/patient lies within the CBCT field of view (FOV), thereby obviating radiation-based scouts. Optical depth technologies (stereo, structured light, LiDAR) are attractive in this role due to their portability, low cost, and non-ionising nature. Among these, stereo cameras such as the RealSense D435 offer a favourable trade-off between performance and ease of integration; however, their use within the strict geometrical and operational constraints of CBCT remains underexplored and is the focus here.

The 3D acquisition system (Figure 4.10) was realised by mounting an Intel RealSense D435 on the gantry of a See Factor CT3 CBCT scanner (Imaginalis Srl, Sesto Fiorentino, Florence, Italy) using industrial hook-and-loop fasteners (Velcro), providing mechanical stability without modifying the CBCT hardware. Based on manufacturer documentation, the scanner's geometric envelope and reference coordinates were known a priori. Combined with the measured mount offset, these specifications allowed us to compute the D435 pose with respect to the X-ray focal point (the origin of the radiation beam), establishing a fixed transform from the camera to the CBCT coordinate frame.

During a simulated imaging workflow, the gantry was rotated around the patient table while the depth camera continuously acquired pointclouds from multiple angles. A ChArUco marker board served as a fiducial for multi-view registration, exploiting its robust detection and subpixel corner localisation. After acquisition, the pointclouds were processed into a single global model through the following steps:

- **Depth filtering:** a lower bound of 28 cm removed near-field artefacts; an upper bound excluded background beyond the table (empirically set to include the full table extent while removing the floor and distant structures).
- **Frame validation:** only frames with reliable ChArUco detection in the RGB image were retained, ensuring that each selected pointcloud could be placed in the global frame via a known pose.

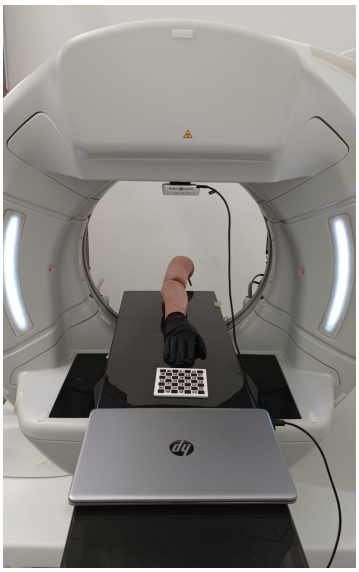


Figure 4.10: Optical scouting setup: Intel RealSense D435 mounted on the See Factor CT3 gantry; a ChArUco board provides a fiducial reference for multi-view alignment.

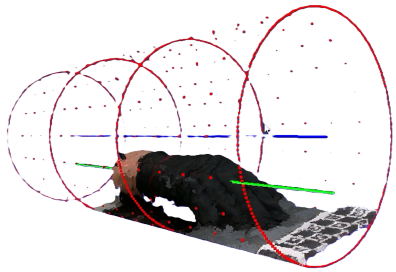
- **Pose consistency (temporal) check:** for each frame, the estimated pose was compared with those of its temporal neighbours. When three consecutive poses were mutually consistent (small relative deviations), their transforms were averaged and the corresponding pointclouds were merged, yielding a temporally stabilised representation (Figure 4.11(a)).
- **Volume cropping and visualisation metadata:** the fused pointcloud was cropped to the known CBCT acquisition volume, removing geometry outside the FOV and highlighting the region physically covered by the scan. To improve interpretability, we overlaid: (i) the scanner isocentre, meaning the centre of the acquisition volume; (ii) the object isocentre, meaning the vertical midpoint of the reconstructed object, used as a practical alignment proxy; and (iii) the stack configuration: the See Factor CT3 supports one to seven longitudinal stacks. This allowed the operator to verify containment within a single stack or distribution across multiple stacks (Figure 4.11(b)).

We assessed spatial accuracy using two block-shaped phantoms (wood and plastic). Each phantom was placed on the CBCT table and imaged in static conditions by both the CBCT system and the gantry-mounted D435. From the CBCT volume, four vertex positions were extracted manually using the scanner's visualisation software, yielding distances from the X-ray focal point. The same vertices were located in the global RGB–D pointcloud produced by our pipeline. Figure 4.12 shows the phantoms and the labelled vertices (A–D).

For each vertex, we compared the distance to the focal point measured in CBCT (reference) versus the optical reconstruction, obtaining an absolute error. Ten repeated measurements were acquired per vertex to characterise variability. Tables 4.1–4.4 report per-trial absolute errors, per-vertex mean ± SD, per-phantom mean ± SD, and the global summary.

These results indicate millimetric spatial fidelity for the proposed optical scouting system: absolute errors were consistently below 1.3 mm, with standard deviations well under 0.3 mm. Considering the low cost and non-ionising nature of the sensing modality, this level of precision is notable. The ability to confirm containment within the CBCT acquisition volume with millimetric accuracy, without scout views, supports the potential clinical utility of the approach.

(a) Global pointcloud after filtering and temporal stabilisation, prior to cropping.



(b) Cropped pointcloud within the CBCT FOV. Red circles: CBCT stacks; blue line: scanner isocentre; green line: object isocentre.

Figure 4.11: Non-ionising optical scouting: reconstruction and field-of-view check.

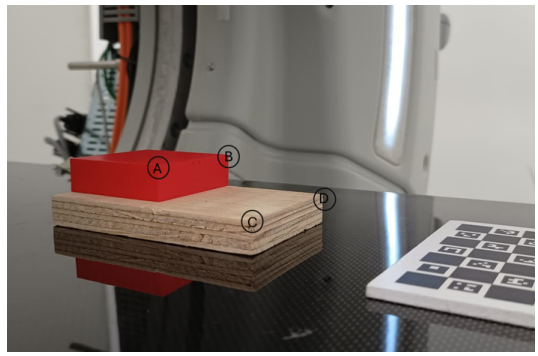


Figure 4.12: Block phantoms and tested vertices (A–D) used for accuracy assessment.

Table 4.1: Absolute errors for each vertex (millimetres) across ten trials.

	Test [mm]									
Vertex	*n°1*	*n°2*	*n°3*	*n°4*	*n°5*	*n°6*	*n°7*	*n°8*	*n°9*	*n°10*
A	1.7	1.3	1.2	1.4	1.5	1.0	1.0	1.1	1.3	1.1
B	1.6	1.0	1.0	1.6	1.6	1.1	1.0	1.3	1.2	1.0
C	1.0	1.0	0.8	1.1	0.9	1.0	1.0	1.0	1.0	1.0
D	1.0	0.9	0.9	0.7	1.0	1.0	1.0	1.1	1.0	1.0

Table 4.2: Per-vertex mean ± standard deviation of absolute error expressed in millimetres.

Vertex	**Mean ± SD [mm]**
A	1.3 ± 0.2
B	1.2 ± 0.3
C	1.0 ± 0.06
D	1.0 ± 0.1

Table 4.3: Per-phantom mean ± standard deviation of absolute error expressed in millimetres.

Phantom	**Mean ± SD [mm]**
Red box	1.3 ± 0.2
Wooden box	1.0 ± 0.08

Table 4.4: Global mean ± standard deviation across all vertices (millimetres).

All vertices (A–D)	1.1 ± 0.2

In conclusion, a gantry-mounted depth camera, when properly registered to the CBCT frame and paired with a robust multi-view pipeline, can provide a reliable, efficient, and radiation-free alternative for positional scouting. These findings motivate future work on integrating the optical scout into routine positioning workflows to prioritise safety, speed, and automation without sacrificing geometric accuracy.

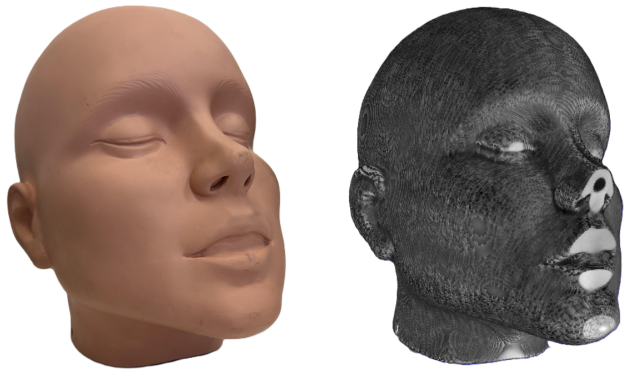
4.4 Texture mapping of CBCT-derived volume

In conventional cone-beam computed tomography (CBCT), internal anatomical structures are captured with high geometric fidelity, but external appearance (texture and colour) is absent. This limits the realism and clinical utility of 3D models derived solely from CBCT, especially where both anatomical accuracy and external appearance are critical (e.g., patient-specific simulation, educational settings, or maxillofacial surgical planning). To address this limitation, we developed and validated a methodology that fuses CBCT-derived volumetric data with photorealistic texture acquired from an RGB–D camera (Intel RealSense D435). External surface data were captured from multiple viewpoints and aligned to the CBCT model through a multi-stage registration pipeline, enabling projection of real texture onto the surface of the CBCT-derived anatomy.

CBCT data were exported in DICOM format and processed to extract a 3D pointcloud. An automated segmentation step (Otsu thresholding per axial slice [45]) was used to isolate the outer contour; where automatic selection failed, manual correction was applied. To obtain a metrically accurate 3D pointcloud, voxel indices were converted to physical coordinates using two DICOM images' parameters: *Pixel Spacing* for the in-plane axes and *Slice Thickness* for the axial axis. In particular:

$$x = X \cdot \frac{PS_x}{1000}, \quad y = Y \cdot \frac{PS_y}{1000}, \quad z = Z \cdot \frac{ST}{1000}, \tag{4.3}$$

where (X, Y, Z) are voxel indices, PS_x, PS_y are the in-plane pixel spacings (mm/pixel), and ST is the slice thickness (mm). This produced a high-resolution pointcloud that accurately represents the scanned object. The human-head phantom and the pointcloud extracted from DICOM images are shown in Figures 4.13(a) and 4.13(b).



(a) Human-head phantom.

(b) 3D pointcloud extracted from DICOM images.

Figure 4.13

To obtain real texture, multiple RGB–D pointclouds were acquired with the D435 from different angles around the phantom. Camera poses were estimated using a 7×7 ChArUco board (as described previously) and OpenCV's pose routines, exploiting known intrinsics and the board geometry. Each RGB–D pointcloud was then transformed into a common coordinate frame and segmented by spatial proximity; artefacts and background were removed via automated clustering and light manual refinement using the MeshLab software. The composite cloud was downsampled with a 1 mm voxel-grid filter (points within 1 mm Euclidean distance averaged), reducing over 10 M points to roughly 70,000. The resulting RGB–D pointcloud is shown in Figure 4.14.

A two-stage registration pipeline aligned the RGB–D pointcloud to the CBCT-derived model:

- **Bounding-box alignment (coarse):** the bounding boxes were translated so their frontal faces coincided, providing an initial spatial correspondence. A detailed description of how this process has been applied has been published in 2024 by Aliani et al. in the article "Optimizing texture representation in 3D medical models using an RGBD camera" [46].

- **2D plane-projection alignment (fine):** both pointclouds were pro-

Figure 4.14: RGB–D pointcloud obtained with the Intel RealSense D435.

Table 4.5: Registration error along each principal axis for different numbers of random points. E_x, E_y, and E_z denote mean ± SD errors (mm).

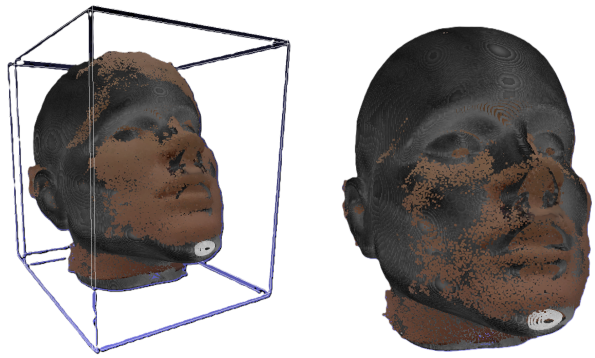
	E_x [mm]	E_y [mm]	E_z [mm]
E_{100}	1.1 ± 1.3	0.5 ± 0.6	0.8 ± 1.0
E_{1000}	1.1 ± 1.2	0.6 ± 0.8	0.9 ± 1.2
E_{10000}	1.1 ± 1.3	0.7 ± 0.8	0.9 ± 1.2
E_{20000}	1.1 ± 1.3	0.7 ± 0.8	0.9 ± 1.2
E_{30000}	1.1 ± 1.3	0.7 ± 0.8	0.9 ± 1.2

jected onto 2D planes defined by randomly sampled rotation vectors. Contours were extracted (concave hull / α-shape), and a distance transform was computed from the CBCT contour. The RGB–D contour was then aligned by minimising the mean distance to this reference via the Nelder–Mead optimiser [47]. The stopping criterion was a mean registration error $E < 1\,\text{mm}$.

The initial bounding-box alignment and the final 2D projection alignment are illustrated in Figures 4.15(a) and 4.15(b).

After registration, texture mapping was performed by associating each CBCT-surface point with its nearest neighbour in the RGB–D cloud and inheriting its colour. A mesh was reconstructed in MeshLab and exported as a textured model (Figure 4.16).

Finally, to assess registration quality, we computed axis-wise average errors on random subsets of the RGB–D points, comparing their aligned coordinates against the CBCT reference. Results are summarised in Table 4.5.



(a) (b)

Figure 4.15: Coarse and fine alignments between the RGB–D pointcloud and the CBCT-derived model. (a): Initial bounding-box alignment; (b): Alignment obtained at the end of the 2D plane-projection algorithm.

Figure 4.16: Textured model obtained by colour transfer from the RGB–D pointcloud.

Errors remained essentially stable as the sample size increased. From $E_{10,000}$ upwards, mean errors stabilised at $\sim$1.1$\pm$1.3 mm (x), 0.7$\pm$0.8 mm (y), and 0.9$\pm$1.2 mm (z). The method proved robust and flexible, achieving millimetre-level alignment using low-cost, non-ionising hardware without reliance on a rigid acquisition setup.

The description of the texturisation process and its potential applications for AR visualisation of the reconstructed 3D model were published by Aliani et al. in 2024 in the article entitled "Realistic Texture Mapping of 3D Medical Models Using RGBD Camera for Mixed Reality Applications" [48].

4.4.1 3D space registration method

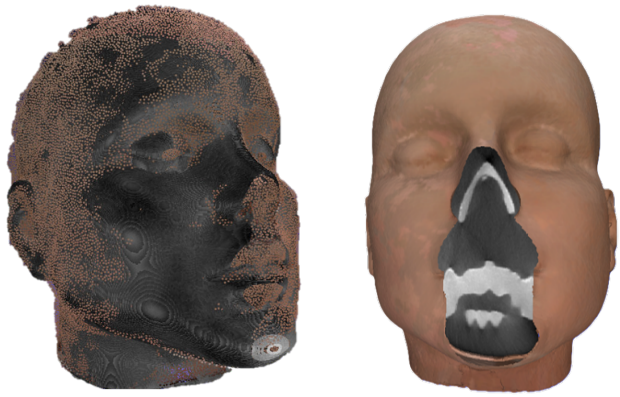
Building on the preceding methodology, ongoing work aims to enhance accuracy and robustness by transitioning from a 2D projection-based alignment to a fully three-dimensional (3D) framework. While the 2D method (silhouette projection + distance-transform minimisation) is effective, it discards information along the projection axis and can introduce ambiguities in complex anatomy.

The updated approach operates directly in 3D. External surface contours derived from CBCT are converted into a volumetric (signed) distance map, and the RGB–D pointcloud is rigidly transformed (rotation + translation) to minimise a cost function defined on the distance field. At each iteration, the transformed points are queried against the 3D distance map and a robust cost (e.g., mean or trimmed mean of distances) is evaluated. Optimisation uses a gradient-free method (Nelder–Mead), suitable for a non-smooth landscape.

This volumetric strategy offers several advantages:

- preserves full 3D structure, reducing ambiguity in concavities and occluded regions;
- removes projection parameters and view-dependent artefacts inherent to 2D methods;
- enables finer-grained optimisation in a consistent reference space.

Preliminary results indicate improved precision in anatomically intricate regions (e.g., nasal cavity, ears, maxillofacial boundaries). A systematic validation comparing 2D vs 3D alignment across phantoms and regions of interest is planned. Early outcomes are shown in Figures 4.17(a) and 4.17(b).



(a) (b)

Figure 4.17: (a): Registered DICOM-extracted and RGB–D pointclouds using the 3D volumetric method. (b): Preliminary result of CBCT volume texturisation.

4.5 3D camera-based simulation of CT imaging on a physical phantom

Computed tomography (CT) is typically taught and practised through two complementary modalities: (i) the interpretation of DICOM image stacks on a workstation and (ii) hands-on training with physical phantoms. In conventional educational settings, however, these two modalities are rarely coupled in a geometrically consistent manner. Trainees usually observe a static phantom in the real world while separately navigating through pre-recorded CT slices on a screen, without an intuitive link between the pose of a slicing plane in space and the corresponding cross-section inside the CT volume. This disconnect makes it difficult to develop spatial intuition—particularly for understanding how changes in scan orientation (including oblique planes) alter the anatomical appearance and the visibility of pathologies.

Bridging this gap is valuable for simulation-based training. A system that allows users to physically manipulate a reference plane around a phantom and immediately visualise the corresponding DICOM slice can emulate key aspects of CT acquisition geometry, support interactive learning, and enable multiple pathological scenarios to be explored without changing the physi-

cal setup. Moreover, if implemented with low-cost sensing and open-source tools, such a framework can be deployed in teaching laboratories and training environments where dedicated tracking systems or high-end simulators are impractical.

This project presents a system that simulates computed tomography (CT) imaging on a physical phantom using an Intel RealSense D435 RGB-D camera. The method combines real-time computer vision with medical image registration to enable interactive visualisation of DICOM slices corresponding to arbitrary planes defined by the pose of a fiducial marker. Such a framework is valuable for medical training, where multiple pathological scenarios can be explored with a single physical setup.

The approach comprises two main components:

- **Real-time pose estimation with a ChArUco marker:** A software pipeline acquires and processes the RGB stream from the D435. Using ChArUco detection and pose estimation, the system computes the six degrees of freedom (6-DoF) of the marker with respect to the camera and updates this pose continuously as the marker moves.

- **DICOM-to-phantom registration via point-based transformation:** To align the phantom's physical coordinate frame with the DICOM CT volume, a rigid registration is computed from corresponding landmarks. Three (or more) anatomically meaningful landmarks are selected within the DICOM volume; the same landmarks are localised on the physical phantom and their 3D positions measured with the ChArUco marker. From these point pairs, a rigid transformation $\mathbf{T}_{\text{ChArUco}\to\text{DICOM}} \in SE(3)$ (rotation $\mathbf{R}$ and translation $\mathbf{t}$; unit scale) is estimated, defining the mapping from the ChArUco/camera frame to the DICOM space.

Once this transformation is established, any new pose of the ChArUco board measured by the camera can be mapped into the DICOM frame. The marker plane then defines an arbitrary slicing plane within the CT volume; the corresponding cross-section is extracted and rendered in real time. An example of the application during the scanning phase is shown in Figure 4.18.

By moving and rotating the marker around the phantom, users visualise cross-sectional slices of the DICOM volume in real time—effectively simulating CT acquisitions along arbitrary (including oblique) planes. Different pathological scenarios can be simulated by simply loading different

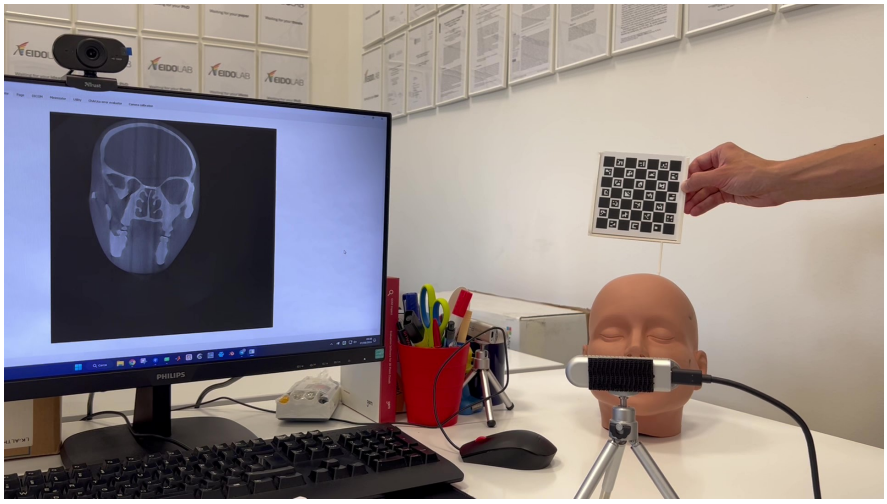


Figure 4.18: Example use with a human-head phantom. The system interactively reslices a DICOM CT dataset acquired on a CT-grade head phantom while the user moves and orients a ChArUco board in space.

DICOM datasets while keeping the same physical phantom. This functionality provides an intuitive and flexible platform for training. It could help learners understand how scan orientation affects anatomical depiction and how pathologies appear under varying slice planes. The system demonstrates that consumer-grade RGB–D hardware, coupled with open-source vision tools, can deliver effective, low-cost simulation for medical education.

4.6 6D pose estimation of an ultrasound probe using neural networks

Reliable localisation of medical instruments in three-dimensional space is critical to image-guided interventions and augmented reality (AR)-assisted procedures. Traditional tracking solutions (optical or electromagnetic) are effective but often costly, complex to set up, and sensitive to occlusions or environmental interference. Recent advances in deep learning and computer vision make markerless tracking a promising alternative—especially when paired with depth-sensing cameras. This section presents the development

and evaluation of a prototype for real-time 6-DoF pose estimation of an ultrasound probe using the *FoundationPose* neural network and RGB–D images acquired with an Intel RealSense D435. The aim was to assess feasibility and accuracy under controlled conditions, with a view to future integration into AR platforms for clinical use.

FoundationPose [49] estimates the full six degrees of freedom (position and orientation) of an object from paired RGB and depth images. Initialisation provides a binary mask for the object of interest in the first frame; thereafter, the network tracks the object in real time without further manual input. The model was chosen for its demonstrated generalisation and modest hardware requirements.

The system was implemented in C++ using Qt Creator (v5.12.6) and the RealSense SDK. An Intel RealSense D435 captured synchronised RGB and depth frames of a clinically relevant probe (Esaote LA533, Italy). RGB and depth were recorded at 1920×1080 and 1280×720, respectively. A binary mask was generated from depth via double-thresholding to isolate the probe from the background.

The tracking pipeline comprised three phases:

- RGB–D frame acquisition;
- binary-mask generation for the initial frame;
- continuous pose estimation with FoundationPose.

Once initialised, the network returned translation and rotation at each frame, updating the 6-DoF pose via incremental refinements from new image data.

Performance was tested under controlled conditions using three known spatial positions (A, B, C) and three yaw rotations (20°, 45°, 60°) about the y-axis. For each position, five repeated measurements were acquired. Translational error was computed from relative distances between positions; rotational error was measured with a goniometer-aligned set-up. Figure 4.19 illustrates the evaluation concept.

The system showed high distance accuracy, with mean translational errors consistently below 6 mm and standard deviations under 1 mm. Rotational accuracy was also satisfactory, with mean errors $< 1°$ across configurations. Sensitivity to initialisation was minimal for translation and modest for rotation, where slightly higher variability in standard deviation was ob-

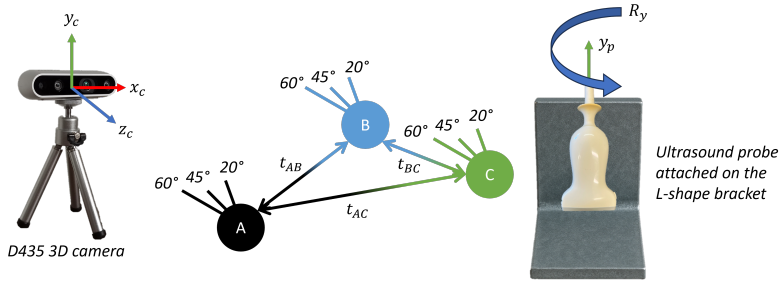


Figure 4.19: Conceptual diagram of the 6-DoF pose-estimation evaluation. Positions A, B, and C were placed at known distances from the camera; at each position, three yaw rotations (20°, 45°, 60°) were tested.

Table 4.6: Summary of mean pose-estimation errors. Distances are reported as mean estimated separations (± SD) compared with ground truth (in mm); rotations report mean estimated yaw (± SD) (in degrees).

Distances	Ground truth [mm]	Mean estimated distance [mm]
t_{AB}	173	171 ± 1
t_{AC}	206	202 ± 2
t_{BC}	176	173 ± 2

Rotations	Ground truth [°]	Mean estimated rotation [°]
ΔR_y	20	19.7 ± 0.9
ΔR_y	45	45.2 ± 1.1
ΔR_y	60	60.2 ± 1.1

served. Table 4.6 summarises mean pose-estimation results while Figure 4.20 shows a screenshot acquired during our test.

These results indicate that FoundationPose can deliver reliable 6-DoF pose estimates of an ultrasound probe in static, controlled environments, with sub-centimetre translational accuracy and sub-degree rotational accuracy. The higher variability in rotation suggests headroom for refinement where angular precision is critical.

This study (published in the IEEE Engineering in Medicine and Biology Society 2025 conference proceedings by Aliani et al. [50]) focused on static pose estimation; dynamic, continuous tracking during probe manipulation remains to be investigated. The setup also covered a limited spatial range. Future work will target dynamic tracking, broader spatial configurations, and

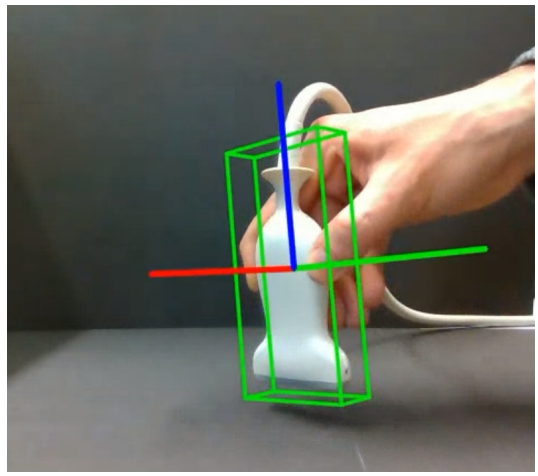


Figure 4.20: Example of application of the FoundationPose neural network on the ultrasound probe. The green box represents the bounding box of the tracked object, while the three coloured axes (green, blue and red) represent the x, y and z axes of the tracked object.

integration with AR headsets (e.g., Microsoft HoloLens 2). By leveraging onboard RGB–D sensing, the probe pose could be visualised in situ within an AR interface, providing clinicians with intuitive, real-time feedback to enhance guidance, training, and telecollaboration.

Chapter synthesis and concluding remarks

This chapter has presented a coherent and progressive line of research on optical 3D imaging systems in biomedical contexts, moving from foundational sensing choices to validated geometric integration and, ultimately, to workflow-facing prototypes. Although the projects were introduced as distinct studies, they are unified by a consistent methodological agenda: first, to characterise depth-sensing modalities under constraints that are typical of clinical environments (challenging materials, illumination variability, and integration constraints); second, to establish reliable registration and pose-estimation primitives as the enabling layer for any downstream use; and third, to embed these primitives into practical applications in which non-ionising, low-cost 3D perception can improve safety, realism, or usability.

A first unifying outcome is the explicit treatment of *sensor modality* as a clinical design decision rather than a purely technical preference. The comparative evaluation of stereo (RealSense D435) versus LiDAR (RealSense L515) under reflective and retroreflective conditions makes a central point that recurs throughout the chapter: in medical environments, nominal precision is insufficient if the sensing mechanism can fail abruptly under plausible operating conditions. In particular, depth integrity and coverage are often more important than best-case accuracy, because downstream pipelines (registration, tracking, and geometric checks) typically degrade catastrophically when depth becomes sparse or inconsistent. The pragmatic selection of a stereo RGB–D camera as the sensing backbone for subsequent work follows directly from this principle, and it sets the tone for the remaining projects: the goal is not to maximise laboratory performance, but to obtain predictable behaviour under realistic constraints.

The second through-line concerns *registration and pose estimation* as the chapter's core enabling technologies. The progression from geometry-driven detection of spherical fiducials (Hough-based), to texture-dependent keypoint matching (SIFT-based), and finally to indexed, subpixel fiducials (ChArUco) can be read as an increasingly deliberate attempt to control uncertainty at the point where it matters most: the rigid transformation that anchors an acquisition to a clinically meaningful reference frame. Importantly, this progression is not framed as an abstract comparison of algorithms; rather, it is grounded in the specific failure modes that arise in biomedical scenes, where surfaces may be weakly textured, visually uniform, reflective, deformable, or partially occluded. The quantitative characterisation of ChArUco pose quality as a function of viewing angle and working distance further consolidates this backbone by turning a widely used tool into a constrained, "deployable" component with actionable operating recommendations—shifting it from a convenient implementation detail to a validated design choice.

Once the sensing modality and the registration primitive are stabilised, the chapter demonstrates how they can be repurposed across heterogeneous biomedical tasks without re-inventing the spatial pipeline each time. The non ionising optical scouting project exemplifies a safety-driven motivation: a gantry-mounted RGB–D camera, registered to scanner geometry and paired with multi-view fusion and temporal consistency checks, can support millimetre level field-of-view containment assessment without ionising scout

views. The CBCT texturisation pipeline then reuses the same geometric integration logic to address a different need—photorealistic appearance—where accurate alignment becomes the enabling condition for credible texture transfer and mixed-reality-ready 3D models, and where the methodological transition from projection-based alignment towards volumetric optimisation reflects a broader trend of reducing ambiguity by retaining more geometry in the optimisation space. The CT simulation framework further extends the same "spatial anchoring" principle to training: mapping the pose of a physical reference plane (tracked via ChArUco) into DICOM space creates an intuitive coupling between real-world manipulation and volumetric navigation, effectively reframing CT interpretation as an interactive, spatial experience. Finally, the probe-tracking study with a foundation model (FoundationPose) illustrates how the same RGB–D sensing stack can support markerless 6-DoF localisation of clinical instruments with promising accuracy under controlled conditions—an important step toward instrument-aware perception without dedicated optical or electromagnetic tracking hardware.

Taken together, the projects support a unified contribution: they show how consumer-grade optical 3D sensors can be turned into clinically interpretable spatial subsystems by combining careful modality selection, robust reference-frame definition, and experimentally validated registration/pose estimation. Beyond individual results, the chapter articulates a transferable engineering pattern: when the clinical task can be formulated geometrically (e.g., containment in a scanner volume, alignment of external appearance to radiological geometry, linkage of physical planes to volumetric slices, or localisation of an instrument in 3D), a small set of well-characterised spatial primitives can be reused across applications—provided that their operating envelope and failure modes are explicitly understood.

Limitations and forward-looking perspective. The chapter also clarifies where the current approach is strongest and where it remains constrained, and these constraints naturally motivate the most relevant next steps. First, many of the most repeatable results rely on explicit geometric constraints (fiducials, calibrated rigs, controlled viewpoints). While methodologically justified—and often essential in visually adverse biomedical scenes—these strategies introduce operational dependencies that matter for deployment: line-of-sight requirements, target placement and sterility considerations, additional setup time, and sensitivity to occlusion or suboptimal viewing ge-

ometry. Second, approaches that reduce explicit instrumentation (feature-based or learning-based methods) shift the burden onto scene appearance and dynamics: texture scarcity, specularity, clutter, and motion can destabilise correspondence and produce intermittent gross errors even when average performance is favourable, which is particularly problematic in clinical workflows where reliability and recoverability dominate perceived usability. Third, several prototypes were intentionally scoped as feasibility demonstrations; consequently, aspects such as automated quality control (knowing when not to trust an estimate), long-term calibration stability, and robustness under operator-driven variability were not exhaustively addressed across all pipelines.

Forward-looking work should therefore focus less on adding standalone features and more on *operationalising* optical 3D imaging as a dependable subsystem. A concrete priority is the systematic integration of quality gates and uncertainty awareness (confidence measures, temporal consistency checks, outlier detection, and safe fall-backs) so that occasional failures are detected early and handled predictably rather than propagating silently. In parallel, the fiducial burden should be reduced without sacrificing repeatability, for example through hybrid schemes in which fiducials provide intermittent re-anchoring while short-term tracking is maintained by markerless cues, or through less intrusive target designs and placement optimisation under realistic occlusion constraints. Validation should be extended from static or controlled acquisitions to dynamic, clinically realistic sequences where motion, occlusions, and workflow interruptions are present, and where usability-relevant endpoints (setup time, recovery behaviour, inter-operator variability) can be quantified. Finally, because several outputs produced in this chapter are inherently compatible with mixed-reality visualisation (textured models, tracked slicing planes, instrument pose), standardising coordinate frames, latency handling, and update rates would strengthen the bridge to the subsequent AR chapters and enable end-to-end spatial coherence from sensing to in-situ visualisation.

Chapter 5

Augmented reality projects

Chapter scope, rationale, and structure

This chapter presents the augmented reality (AR) line of research developed in this dissertation. Rather than treating AR as a generic "visual gadget", the guiding premise across all projects is that AR becomes clinically meaningful when it (i) externalises complex three-dimensional information into the user's workspace, (ii) preserves interaction fluency under real constraints (latency, compute, sterility, ergonomics), and (iii) connects to real clinical artefacts—images, patients, workflows, and multidisciplinary collaboration.

The projects described here form a coherent progression around a single research question: *How can head-mounted AR be engineered as a practical, reusable visualisation-and-interaction layer for medical data that supports both pre-operative reasoning and procedure-adjacent workflows?* Throughout the chapter, the "application" is never the final objective per se; each application serves as an instrument to stress-test and refine a small set of reusable AR building blocks.

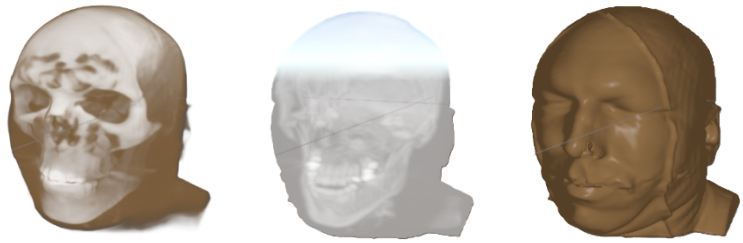
The chapter is organised as a roadmap of augmented-reality projects carried out during the PhD, designed to present a progressive, unified line of research. The first section establishes the enabling AR platform by implementing interactive volumetric visualisation on HoloLens 2 (rendering modes, transfer functions, and slicing), and then extends it with server-driven dataset access and patient-centred workflow modules. The second section addresses a core prerequisite for clinically grounded AR, namely spatial anchoring, by developing a QR-code–based registration strategy to align

CT-derived models with physical references. The third section broadens the interaction paradigm beyond rigid manipulation by introducing a real-time soft-tissue deformation model for digital liver palpation. The fourth section then moves to procedure-adjacent use, treating AR as a human-factors intervention in interventional ultrasound: it first defines a low-latency streaming pipeline and subsequently evaluates two representative tasks (breast biopsy and ultrasound-guided cannulation). The next section integrates AI-based organ and pathology segmentation into the AR workflow by coupling 3D Slicer and Unity via OpenIGTLink, enabling end-to-end generation and holographic inspection of co-registered parenchyma, vessels, and tumours for hepatobiliary planning. Finally, the last sections generalise the platform along two complementary dimensions: dynamic data (4D volume visualisation) and collaborative access (a web-based viewer synchronised with real-time AR co-manipulation). Overall, the organisation deliberately traces an incremental transition from foundational AR capabilities to clinically motivated prototypes and, ultimately, to scalable extensions that support richer data and multi-user, cross-platform interaction.

5.1 Medical 3D volumes visualisation

The growing integration of augmented reality (AR) in medicine—especially in surgery—has enabled innovative visualisation tools for preoperative planning and intraoperative guidance. Within this context, the present project developed an AR application for interactive visualisation and manipulation of three-dimensional (3D) medical data. The objective was to render volumetric medical images as holograms in the real environment using Microsoft HoloLens 2, allowing clinicians to explore and interact with anatomy in a spatially coherent manner.

This work adapts and extends the open-source *UnityVolumeRendering* plugin [51]—originally designed for standard displays—to a fully immersive AR setting. The application, built with Unity and the Mixed Reality Toolkit (MRTK), adopts a dual architecture: a desktop "Host" application and an AR "Client" on HoloLens 2. Holographic Remoting offloads computation from the headset to ensure smooth rendering and interaction. The system supports DICOM, NIfTI, and RAW volumes, with multiple rendering techniques: Direct Volume Rendering (DVR), Maximum Intensity Projection (MIP), and Isosurface Rendering (IR). Figure 5.1 shows examples on a hu-



(a) Direct Volume Rendering (b) Maximum Intensity Projection (c) Isosurface Rendering

Figure 5.1: Examples of the supported rendering techniques applied to a head CT volume.

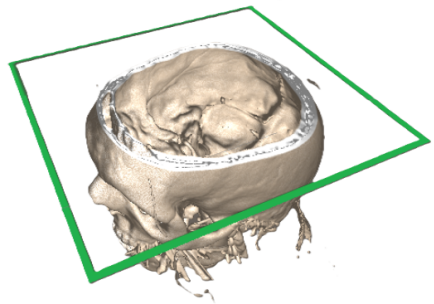


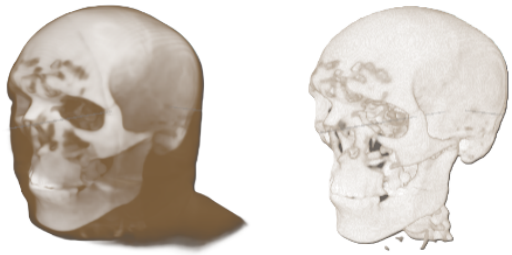
Figure 5.2: Planar cross-sectioning of the rendered volume.

man head dataset; each technique emphasises different anatomical features.

Rendering can be refined via one- and two-dimensional transfer functions that map voxel intensity and gradient to opacity and colour, enabling targeted emphasis of regions such as bone or soft tissue. Internal structures are explored using interactive cross-section tools: users define cutting planes or volumes (cubes and spheres) and adjust position, orientation, and scale, with *inclusive* and *exclusive* slicing modes. An example of a planar cut is shown in Figure 5.2.

A slice-rendering window facilitates precise analysis of 2D sections, with tools for pixel-intensity readout and distance measurement. Figure 5.3 illustrates the effect of adjusting the minimum intensity threshold.

User interaction is provided via an MRTK-based hand menu rendered on



(a) Original volume.

(b) Volume after increasing the minimum intensity threshold.

Figure 5.3: Effect of intensity thresholding on the rendered volume.

the user's palm (Figure 5.4(a)), offering an intuitive interface for all controls. Through this interface, users can enable cross-section objects (Figure 5.4(b)), switch rendering modes, adjust visibility, and more, using natural hand gestures.

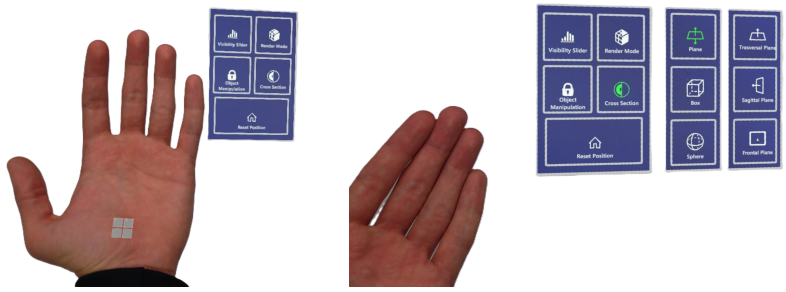
The desktop Host complements the AR client by managing dataset loading and providing an extended control panel. It offers a graphical interface to configure rendering parameters (Figure 5.5(a)), edit transfer functions (Figure 5.5(b)), and manipulate slicing planes. Interactions are mirrored to the AR scene, enabling collaborative workflows between multiple users or between a surgeon and technical staff.

The outcomes demonstrate the feasibility and utility of AR-based volumetric visualisation for medical use. The system enables immersive, interactive inspection of complex anatomy, supporting spatial understanding that is difficult to achieve with 2D images. Figure 5.6 shows a user moving a cutting plane with natural hand gestures.

In conclusion, the developed AR system illustrates how immersive technologies can enhance medical image interpretation and surgical planning.

5.1.1 Server-based volume management and patient workflow integration

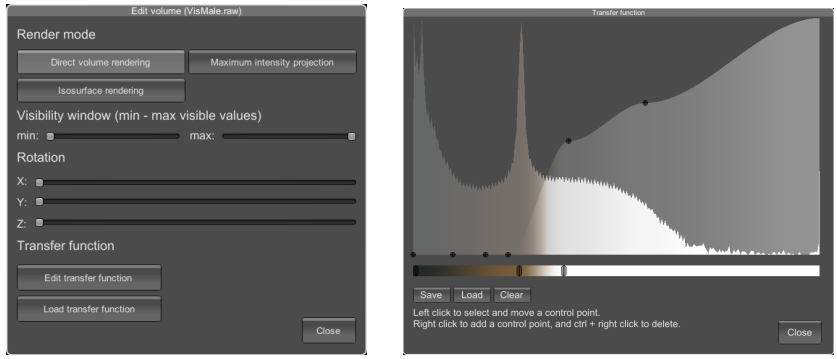
Building on the previous section, this development extends the AR visualisation system with server–client communication to enable dynamic retrieval



(a) Hand menu rendered on the user's palm.

(b) Hand-menu controls for enabling different types of cutting planes.

Figure 5.4: In-situ user interface for AR volume manipulation.



(a) Desktop menu for volume configuration

(b) Desktop editor for transfer functions

Figure 5.5: Host-side controls mirrored to the AR client.

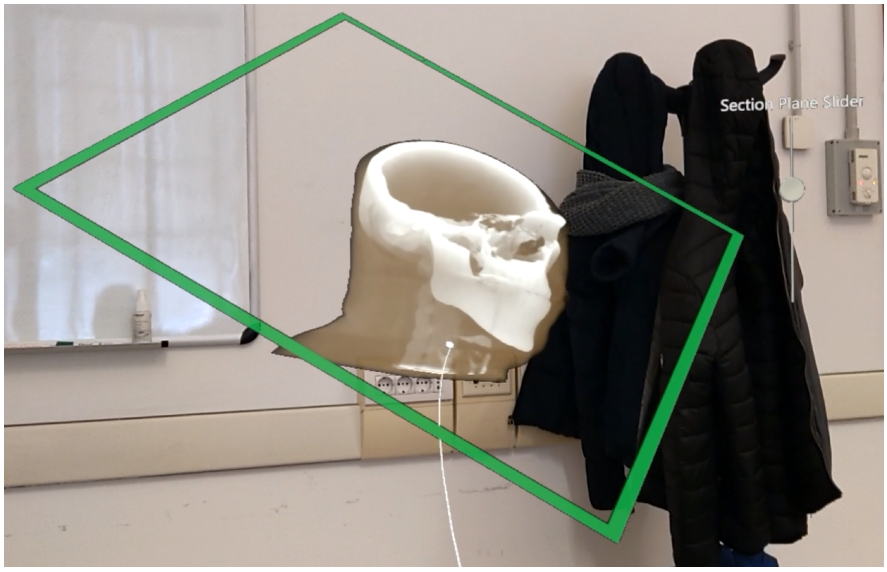


Figure 5.6: User interaction with a cutting plane via hand gestures.

of remotely stored medical data. The goal was to overcome local-only access and provide a flexible, scalable solution for real-time interaction with centralised datasets. The architecture adds a dedicated communication layer, called `ServerCommunicationManager`, which issues HTTP requests, receives JSON responses, and deserialises them into a navigable directory tree.

An intuitive hand-based UI, shown in Figure 5.7 and implemented with Unity and MRTK, lets users browse the server in AR like a conventional file explorer. The interface adapts dynamically to directory contents and supports context-sensitive actions by file type.

Files are downloaded asynchronously and handled according to their type:

- **DICOM, RAW and NIfTI** files are passed to dedicated downloaders and instantiated as interactive 3D volumes.
- **Text** files (e.g., clinical notes) are displayed on holographic panels that are scrollable and relocatable in 3D, with optional gaze-tracking to maintain visibility during user movement (Figure 5.8).

Users can switch seamlessly between the server browser and the volume-manipulation interface via a hand-menu toggle. Once a volume is loaded,

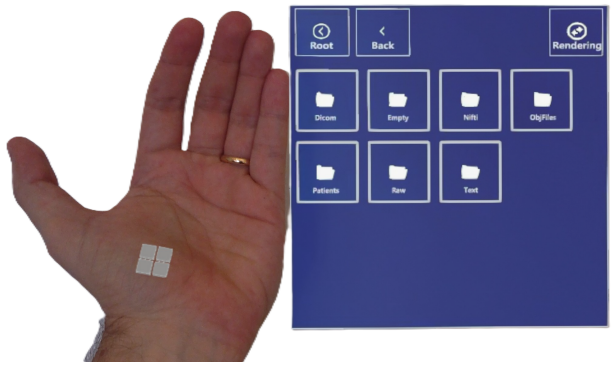


Figure 5.7: Hand menu for browsing remote folders within the AR scene.

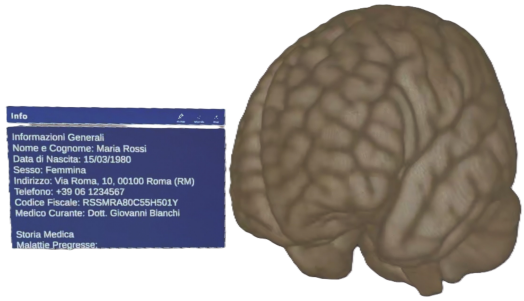
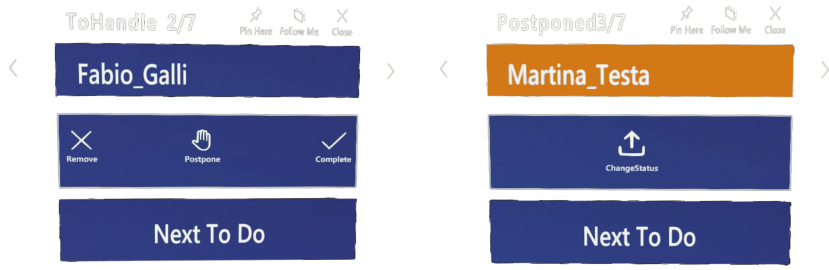


Figure 5.8: Visualising, reading, and scrolling clinical notes in AR.



(a) AR menu for patient-list management (b) Menu state after updating a patient's status

Figure 5.9: Patient-centred workflow integrated into the AR interface.

the full interaction toolkit is available: visibility sliders, rendering-mode selection, cross-sectional exploration, and spatial manipulation.

A further contribution is a patient-specific workflow module. A custom file extension (`.plst`) identifies patient lists. When such a file is detected, the system parses and displays patient names with associated metadata (relative path, clinical status). Four statuses are currently defined: *Pending*, *Managed*, *Postponed*, and *Discarded*. As shown in Figures 5.9(a) and 5.9(b), users can update status in real time via a dedicated menu and jump directly to the next actionable record.

This extension transforms the platform from a static viewer into a dynamic, patient-centred data tool. Direct interaction with server-hosted volumes, combined with integrated patient tracking, makes the solution promising for intraoperative use and clinical decision support.

5.1.2 Optimisation of the transfer function

Building on the volumetric rendering capabilities above, this project refined transfer functions to enhance the appearance of 3D medical datasets. The aim was to replace default greyscale mappings with more expressive, perceptually intuitive colour schemes that accentuate anatomical features.

A systematic tuning of the one-dimensional transfer function—mapping voxel intensity to colour and opacity—was performed on an abdominal CT volume containing heterogeneous structures (bone, soft tissues, organs). By adjusting gradient and colour-mapping curves (see Figure 5.5(b)), regions of

interest were highlighted with greater clarity and aesthetic appeal.

Colour palettes were selected for perceptual quality, seeking results that appear more vivid and realistic than monochrome renderings. This optimisation was conducted without clinical supervision and is not intended to provide medically validated views; perceptual clarity, colour harmony, and structural emphasis were prioritised over diagnostic accuracy. The resulting transfer functions were evaluated qualitatively for improved differentiation between regions and natural integration in AR. Figures 5.10(a) and 5.10(b) respectively depict clipping-plane interaction with the volume prior to and following the application of the optimised transfer function.

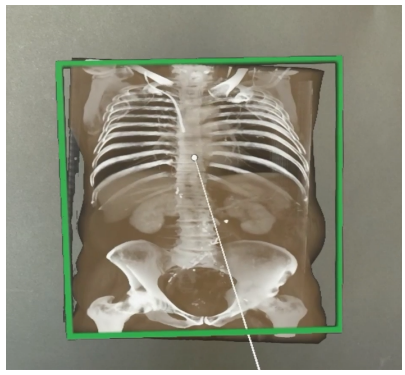
Although not clinically validated, these perceptually tuned mappings enhance user experience during volumetric exploration. They foster engaging, informative interactions—especially in educational or preoperative rehearsal scenarios where anatomical realism and recognisability are valuable.

Significance. This project lays the technical and conceptual foundation for the AR contributions that follow. By establishing a robust pipeline from medical images to holographic 3D volumes—with interactive slicing and perceptually tuned colour mappings—it provides the reusable rendering core, interaction patterns, and data pathway (DICOM/RAW/NIfTI → AR) on which subsequent systems build. In this sense, it is the first stone of the AR projects of this thesis: a general, extensible substrate that enables server-driven data access, patient-centred workflows, multimodal overlays, and task-specific AR tools presented in the next sections.

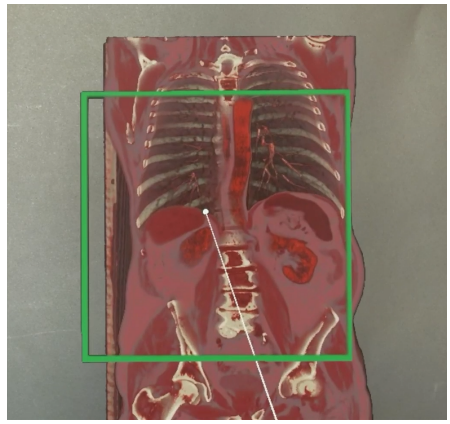
5.2 Volume registration with QR code

In recent years, augmented reality (AR) has attracted significant attention in medicine for its ability to superimpose three-dimensional virtual anatomy onto the physical world. This capability enhances visualisation and interaction in surgical planning and navigation. A pivotal challenge for AR systems such as Microsoft HoloLens 2 is accurate registration between virtual reconstructions derived from medical imaging (e.g., CT) and the corresponding physical anatomy. Registration approaches are commonly categorised by hardware set-up and interaction paradigm:

- **Manual methods:** rely on user input to align the hologram to the real anatomy, either through fully manual placement (voice/joystick)



(a)



(b)

Figure 5.10: Visualisation of the same abdominal volume before (a) and post (b) the application of the optimised transfer function.

or fiducial markers on the patient's skin. While straightforward, these methods are imprecise and not robust to motion.

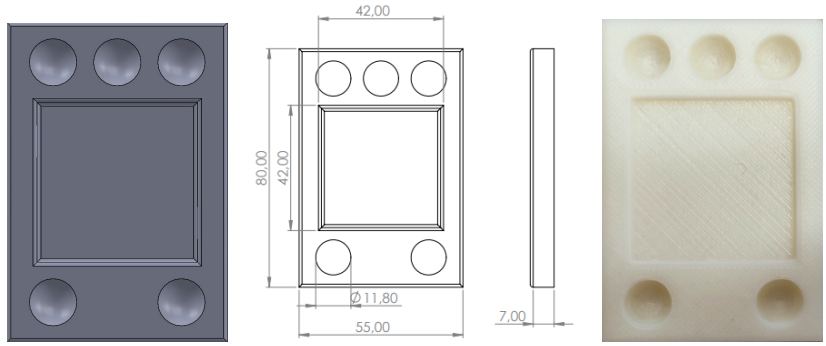
- **Inside-out methods:** use sensors embedded in the headset. Marker-based variants depend on visual tags (e.g., ArUco markers and QR codes); they avoid external trackers but may suffer from recognition instability and latency. Marker-less variants (e.g., Simultaneous Localisation and Mapping, SLAM) exploit environmental mapping or anatomical landmarks, but typically lack surgical-grade precision.
- **Outside-in methods:** use external tracking systems (infrared cameras, electromagnetic sensors). These deliver high accuracy but increase set-up complexity and can be constrained by line-of-sight.

The goal of this project was to develop and qualitatively evaluate an automatic method to register a CT-based digital model with its physical counterpart on HoloLens 2. We implemented an inside-out, marker-based approach that leverages the headset's built-in cameras and uses a QR code as a spatial anchor.

A custom 3D-printed PLA mount was designed in SolidWorks. The mount includes five hemispherical cavities for spherical markers and a central housing for the QR code. Its symmetry and dimensions were optimised to preserve orientation and positioning between CT acquisition and AR registration. A CT scan of a human-heart phantom—equipped with three spherical markers—was acquired at Imaginalis Srl. (Sesto Fiorentino, Florence, Italy) using a Pico3030 scanner. DICOM images were processed in 3D Slicer to extract a segmented OBJ model, which was scaled to Unity's metric convention. Figure 5.11 shows the mount's design and fabrication stages; Figure 5.12 shows the physical phantom and its digitised model.

Registration was implemented in a custom C# component, which identifies marker positions, computes the rigid transform aligning the model to the QR code frame (rotation and translation), and exposes a minimal UI to include the repositioning function. We evaluated three marker-based tracking options:

- **ArUco:** detected ArUco markers but exhibited notable latency and instability on HoloLens 2.
- **Vuforia Engine (an AR computer-vision SDK for image/-marker target detection and tracking):** provided faster detection

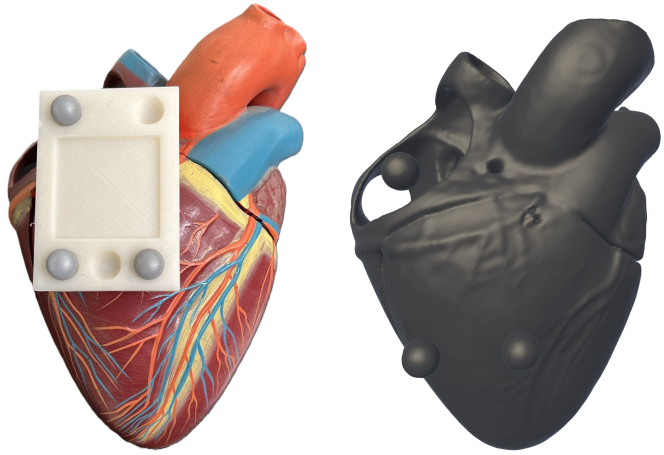


(a) SolidWorks sketch of the PLA mount.

(b) Dimensioned sketch.

(c) 3D-printed mount.

Figure 5.11: Design and fabrication of the PLA mount that houses spherical markers and the central QR code.



(a) Heart phantom with PLA mount and spherical markers.

(b) Digitised model obtained with the Pico3030.

Figure 5.12: Physical phantom and corresponding CT-derived digital model.

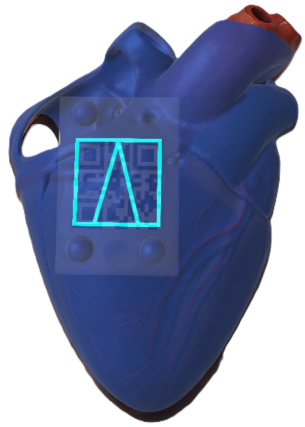


Figure 5.13: Alignment result using QR code–based registration.

but produced inaccurate placement due to unresolved positional offsets and scale inconsistencies.

- **QR code via MRTK extension:** offered fast and accurate registration by recognising a static QR code and aligning the model accordingly. A holographic hand menu supports scene switching and reset of QR detection.

Figure 5.13 shows the alignment achieved with the QR code method.

A clipping plane (Figure 5.14) was added to enable real-time exploration of internal anatomy, enhancing interactive educational and diagnostic use.

The QR code method outperformed ArUco and Vuforia in precision, stability, and usability. Alignment was visually verified by matching virtual spheres to the mount's physical cavities. Unlike Vuforia—which introduced systematic offsets—the QR code approach yielded accurate superimposition with minimal perceptual error. Qualitative comparison indicated that:

- QR code–based registration achieved high precision and stable tracking;
- alignment latency was low and robust under moderate user motion;
- the repositioning mechanism improved usability;
- the 3D-printed mount was adaptable and reusable across scenarios.

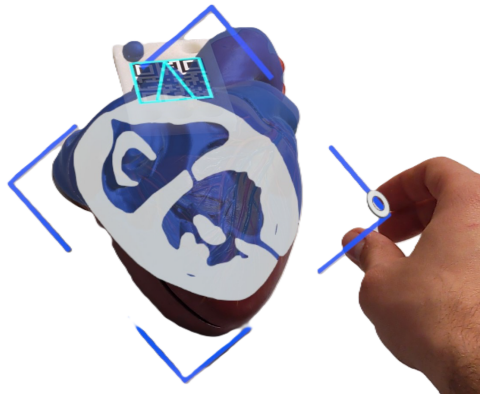


Figure 5.14: Interactive clipping plane used to explore internal structures.

Limitations.

- Registration relies on on-board HoloLens 2 processing, constraining model complexity.
- QR code recognition is performed once per session unless manually reset using the repositioning button in the UI.
- The current module is standalone and should be integrated into a broader AR surgical application.

Future work. Planned enhancements include Holographic Remoting to offload computation while maintaining camera access, detachable marker stands for non-invasive patient use, and objective accuracy assessment with quantitative metrics.

5.3 Digital liver palpation

This section presents the design and development of a mixed-reality system that simulates digital palpation of liver tissue. The project integrates real-time 3D interaction with a segmented liver mesh on Microsoft HoloLens 2, aiming to provide a realistic and intuitive experience suitable for both clinical and educational contexts.

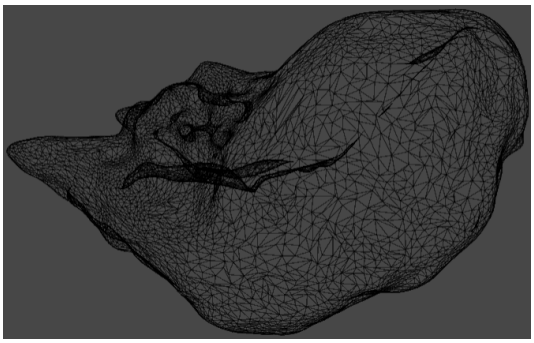
Figure 5.15: Liver mesh obtained from segmentation of CT images.

The core idea is to create an immersive environment in which users visualise and manipulate a holographic liver with their hands. The liver model (Figure 5.15), obtained from CT images segmentation, was imported into Unity and optimised for HoloLens 2. Given device constraints, high-resolution meshes were downsampled to maintain smooth performance without materially compromising anatomical fidelity. Among the tested meshes (Figure 5.16), the medium-accuracy model (Figure 5.16(b)) offered the best balance between visual detail and computational efficiency.

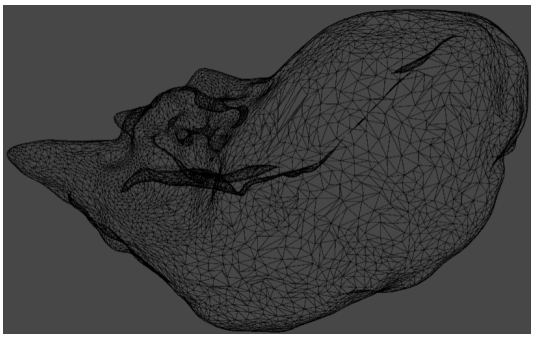
The development environment relies on Unity and the Mixed Reality Toolkit (MRTK), which provides hand-tracking and spatial-mapping primitives. Once deployed to HoloLens 2, the liver mesh becomes an interactive hologram. Hand tracking instantiates ten invisible *capsules*—one per finger tip—that move in real time with the user's hands and act as the interface to the virtual organ. The finger-tip capsules and the base capsule are shown in Figure 5.17 and Figure 5.18.

Contact detection combines spherical colliders with ray-casting. From each capsule, five rays are emitted in different directions; when rays intersect the liver mesh, candidate contact points are computed. The shortest valid ray determines the active contact point, from which a deformation force is derived based on contact proximity and surface orientation. Force magnitude increases linearly as the capsule approaches the surface, up to a fixed maximum.

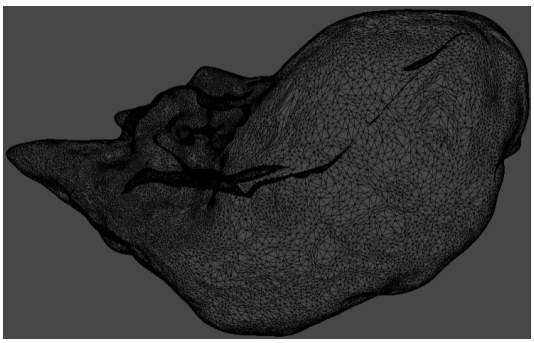
Upon contact, a dedicated script triggers mesh deformation by passing the contact point, direction, and intensity to the deformation module. Deformation is local: only vertices within a radius of influence are displaced, with displacement attenuating with distance. A spring–damper model and a cap on maximum offset approximate soft-tissue behaviour. To better em-



(a) Low-accuracy liver mesh



(b) Medium-accuracy liver mesh



(c) High-accuracy liver mesh

Figure 5.16: Meshes evaluated for AR palpation: quality–performance trade-off.

Figure 5.17: Finger-tip capsules (one per finger).

Figure 5.18: Basic interaction capsule.

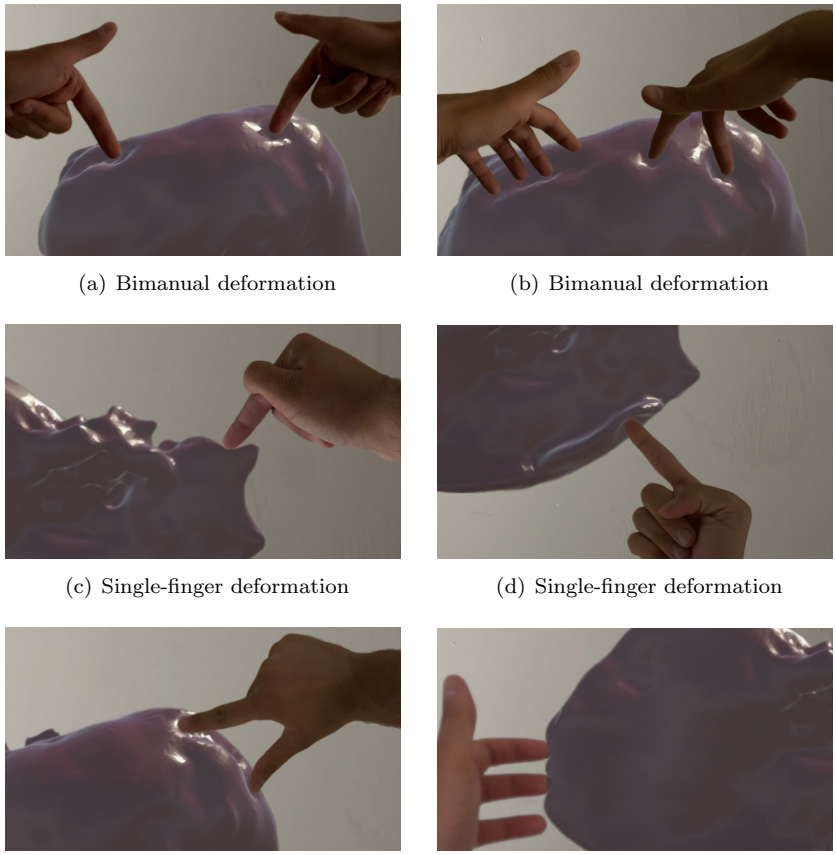
ulate force propagation, a dual-Gaussian radial profile (Eq. 5.1) modulates the influence field, producing a sharp central response that decays smoothly. Vertices are eligible for deformation only within 90% of the profile's support, which improves efficiency and prevents unrealistic distortions.

$$f(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_2^2}} \tag{5.1}$$

The system supports simultaneous multi-finger and bimanual interaction—an advance over single-finger approaches—so multiple regions can deform concurrently, yielding palpation dynamics closer to real practice. To reduce computational load, only affected vertices are updated per frame. A hash map tracks each active vertex and its velocity, which evolves under elastic and damping forces; this targeted update strategy improves responsiveness substantially. Representative interactions are shown in Figure 5.19.

Parameter tuning and calibration. All deformation parameters—spring stiffness and damping, influence radius, maximum offset, and the dual-Gaussian widths (σ_1, σ_2)—were tuned empirically to obtain a visually plausible, smooth response on HoloLens 2 under real-time constraints. This configuration is *not* clinically calibrated. For translation into clinical practice, parameters should be identified against clinician feedback and physical references, ideally using indentation tests or elastography/MRE-derived stiffness maps to fit a local viscoelastic model (e.g., minimising error between simulated and reference force–displacement curves). Lesion simulation would require *spatially varying* properties (e.g., higher local stiffness, altered damping) to reflect tumour heterogeneity and depth, with iterative refinement based on expert palpation ratings. If haptic feedback is added, the same calibration loop should jointly tune visual deformation and haptic gain to maintain perceptual consistency.

Outlook and generalisability. AR with HoloLens 2 opens a broad range of applications, from surgical planning to medical education. For training, the simulation enables risk-free practice of palpation techniques and the potential identification of subsurface anomalies by feel. With the addition of haptics (e.g., a wearable glove), the system could evolve into a comprehensive tool for hands-on learning and pre-operative rehearsal. Although we focus here on liver palpation, the framework is organ-agnostic: extending to other



(a) Bimanual deformation

(b) Bimanual deformation

(c) Single-finger deformation

(d) Single-finger deformation

(e) Multi-finger deformation

(f) Multi-finger deformation

Figure 5.19: Examples of user–mesh interactions and resulting deformations.

body districts (e.g., abdominal wall, breast, thyroid, prostate, musculoskeletal masses) or to tool-mediated palpation in laparoscopic scenarios mainly entails substituting the anatomical mesh and adjusting material parameters and boundary conditions. Pathology-specific simulations can be achieved by embedding focal inclusions with distinct viscoelastic profiles and calibrating them with clinician feedback.

5.4 Augmented reality for interventional ultrasound

Interventional ultrasound (US) is ubiquitous in procedures such as venous cannulation, biopsies, and targeted injections. Yet, a persistent ergonomic limitation remains: operators must repeatedly shift their gaze between the puncture site and the ultrasound monitor, usually positioned on their left or right. This head–eye oscillation fragments attention, complicates hand–eye coordination, and can increase cognitive load—particularly in confined spaces or when the display is suboptimally placed. Moreover, frequent head turns may disrupt a sterile posture and reduce the operator's situational awareness around the needle tip.

This project explores augmented reality as a human–factors intervention: instead of relocating the clinician to the ultrasound image, we relocate the ultrasound image to the clinician. Specifically, the live US feed is rendered in situ on a spatially anchored, holographic display plane within the operator's field of view, using Microsoft HoloLens 2. The hologram's position, orientation and scale can be adjusted with natural hand gestures so that the image can be placed where it is most useful (e.g., just above the insertion line-of-sight, or offset to avoid occluding the patient).

Aims and scope. The aim is not to alter ultrasound acquisition or interpretation per se, but to improve the presentation of the existing signal at the point of action. Concretely, we set out to:

1. Stream the real-time ultrasound image to HoloLens 2 and render it as a stable holographic panel, with low perceived latency;

2. Provide intuitive, sterile-field–compatible interaction (hands/voice) for repositioning and resizing the panel during a procedure;

3. Demonstrate the approach in two representative tasks: (i) breast tumour biopsy and (ii) ultrasound-guided cannulation.

Design considerations. The system was designed around several constraints typical of interventional settings:

- **Ergonomics and attention:** minimise head/eye shifts and preserve continuous line-of-sight to the needle–target trajectory.
- **Latency and stability:** ensure smooth, near–real-time visual feedback and robust anchoring of the holographic panel under operator motion.
- **Sterility and workflow:** support hands-free or gloved-hand interactions; avoid additional hardware in the sterile field.
- **Readability:** maintain sufficient luminance and contrast of the US image in mixed lighting.
- **Non-intrusiveness:** prevent occlusion of critical anatomy and allow rapid repositioning or dismissal of the panel.

Structure of the section. We first describe the video streaming pipeline and AR rendering of the ultrasound feed (§5.4.1). We then present two focused applications: breast biopsy (§5.4.2) and ultrasound-guided cannulation(§5.4.1).

5.4.1 Ultrasound machine-to-AR streaming

A core building block for AR-assisted, ultrasound-guided procedures is the ability to stream the live ultrasound image directly into a head-mounted display. The goal is to reduce head–eye switching between the puncture site and a remote monitor by presenting the same US feed on a spatially anchored holographic panel within the operator's field of view.

Architecture. We implemented a client–server pipeline. On a standard PC, the server captures the US video via an Elgato CamLink 4K HDMI acquisition device and transmits frames over a persistent WebSocket connection. This keeps latency low and throughput stable for fluid visual feedback. On the client side, a Microsoft HoloLens 2 application receives the stream

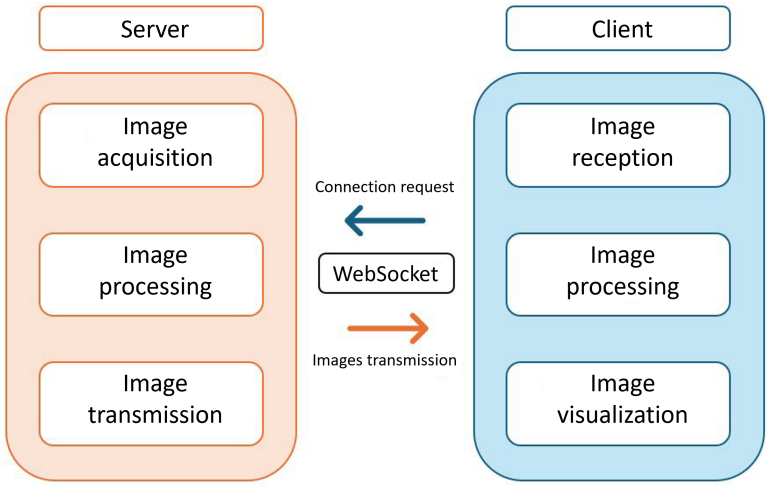


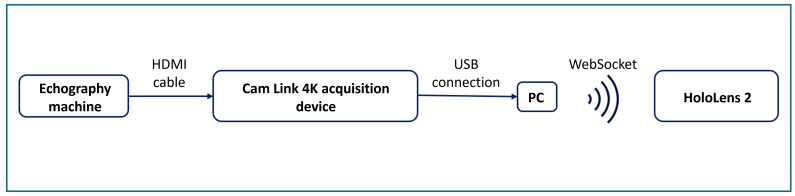
Figure 5.20: Schematic of the client–server pipeline: HDMI capture on the PC, WebSocket streaming, AR rendering on HoloLens 2.

and renders it as a holographic AR "monitor". Figure 5.20 summarises the system behaviour, while Figure 5.21 shows the capture device and overall structure.

Interaction and ergonomics. The virtual monitor (rendering surface) can be freely positioned and rotated with hand gestures, anchored in space, and resized to optimise readability and bandwidth. A server-side region-of-interest tool allows cropping the source image before transmission (Figure 5.22). All controls are exposed through an MRTK hand menu (Figure 5.23), supporting sterile-field–compatible interactions.

Deployment and qualitative feedback. We evaluated the system in two settings at Santa Maria Nuova Hospital (Florence, Italy): (i) simulated needle insertion on meat phantoms (Figures 5.24(a)–5.24(b)) and (ii) non-invasive scanning on volunteers (Figure 5.24(c)). In both scenarios, clinicians reported improved hand–eye coordination and spatial awareness. The ability to reposition the holographic panel—rather than repositioning the operator relative to a fixed monitor—was perceived as intuitive and adaptable to varied room layouts.

(a) Elgato CamLink 4K [52]



(b) High-level system structure

Figure 5.21: Hardware capture and end-to-end streaming set-up.

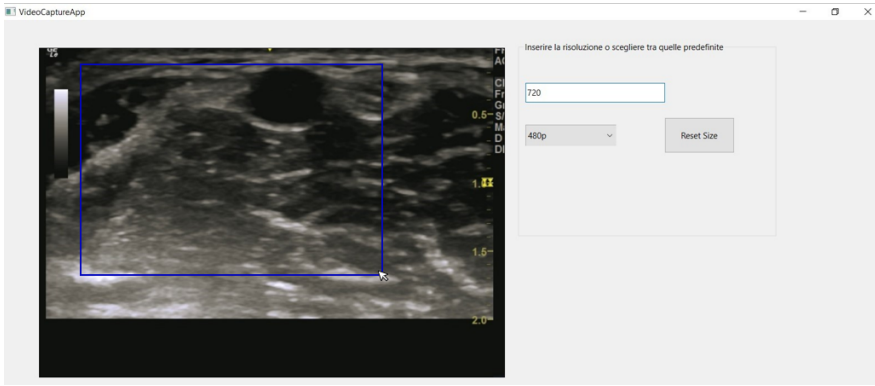


Figure 5.22: Selecting the optimal region of interest on the server to reduce bandwidth and improve readability on HoloLens 2.

Figure 5.23: Hand menu for in-situ controls.

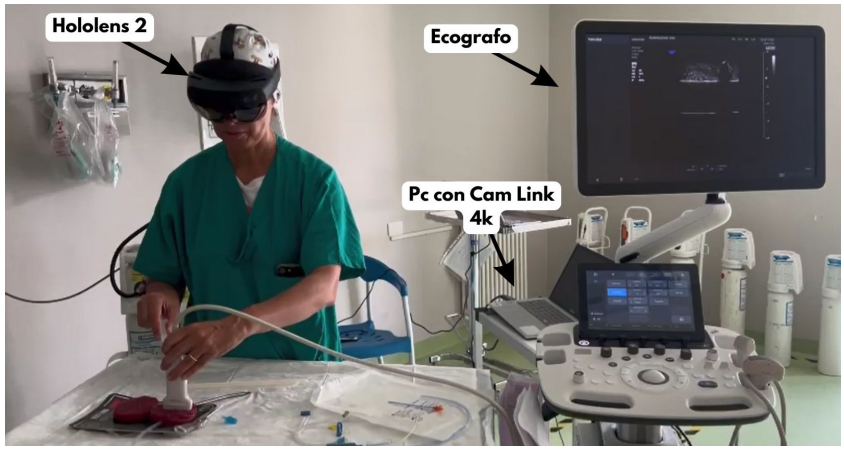
Limitations. Despite positive feedback, hardware constraints of the AR headset can limit performance. High-resolution decoding and AR compositing under motion may induce frame drops, and perceived latency depends on network conditions and encoder settings. These constraints inform the subsequent focus on task-specific layouts and further optimisation.

Summary. This streaming module provides a flexible foundation for AR-assisted ultrasound. It generalises across interventional use cases (e.g., vascular access, biopsies) and serves as a reusable component for the applications detailed in the following subsections.

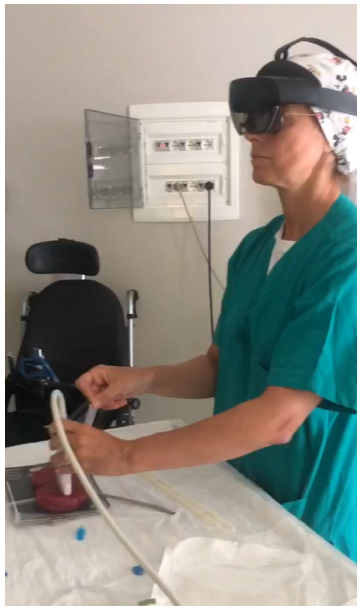
5.4.2 AR-assisted ultrasound-guided breast biopsy

Aligning ultrasound imagery with the operator's natural line of sight can reduce head and gaze shifts between the patient and the ultrasound machine monitor, with potential gains in efficiency, ergonomics, and situational awareness during biopsy. Building on the AR streaming framework described in the preview section, we evaluated a workflow that renders the live B-mode stream as a spatially anchored holographic display, freely positioned and resized by the user.

The imaging pipeline comprised a portable ultrasound unit with HDMI output, an Elgato CamLink 4K capture interface, a laptop for video acqui-



(a) Simulation of needle insertion on a meat phantom



(b) Simulation of needle insertion on a meat phantom

(c) Non-invasive scanning on a volunteer

Figure 5.24: Pilot evaluations: simulated cannulation on phantoms and ultrasound scanning on volunteers.

Figure 5.25: Insertion of a dark olive in the chicken breast.

sition/processing (OpenCV) and transport (WebSocket), and a Microsoft HoloLens 2 client that decodes the stream and presents it on a manipulable quad in the operator's field of view.

Experiments were performed on standardised tissue-mimicking phantoms prepared from chicken breast. Small randomised incisions formed sub-surface pockets into which dark olive fragments were inserted; the dark inclusions enabled unambiguous confirmation of successful sampling within the biopsy needle. A photo acquired during a chicken breast preparation is shown in Figure 5.25.

Data acquisition took place in a breast imaging unit at Careggi University Hospital (Florence, Italy). All participants were *AR-beginner*, with no prior hands-on experience of HoloLens 2 or comparable head-mounted AR systems. The study proceeded in two phases:

- **Phase 1:** eleven staff participants each executed two trials across four operating configurations;
- **Phase 2:** a single operator performed ten repeated trials per configuration to assess repeatability.

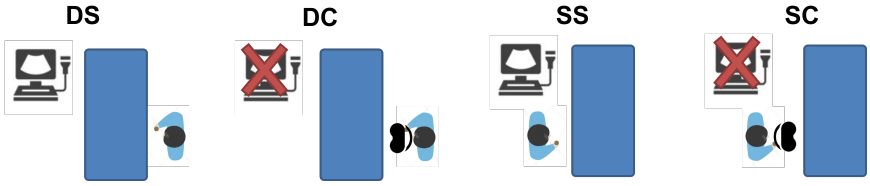


Figure 5.26: Schematic representation of the four configurations used during the biopsy test, respectively: opposite side without AR visor (DS), opposite side with AR visor (DC), same side without AR visor (SS) and same side with AR visor (SC).

The four configurations crossed operator position (same side as the ultrasound machine vs. opposite side) with visualisation modality (conventional monitor vs. HoloLens 2 AR). The same-side, monitor-based setup reflects routine clinical practice and served as the traditional reference condition. A schematic representation of the four configurations is shown in Figure 5.26, while a picture of one of the operators during an AR-guided phase is shown in Figure 5.27.

Two timing endpoints were recorded for each trial:

- **Detection time:** from probe contact to sonographic detection of the target olive;
- **Extraction time:** from detection to successful sampling with the biopsy needle.

Additionally, a structured questionnaire was submitted, aiming to assess perceived usability, visual clarity, freedom to position the image in space, overall satisfaction, and willingness to recommend the AR workflow to colleagues.

Results. Total procedure time was defined as the sum of *detection* and *extraction* for each attempt. In the multi-operator phase (Phase 1; Figure 5.28), although group means remain approximately constant across configurations, there is a tendency for total times to increase when the AR headset is used, irrespective of operator side. In the single-operator phase (Phase 2; Figure 5.29), an evident anomaly emerges: trials performed in the *DC* configuration are markedly longer than all other configurations, with substantially elevated central tendency and dispersion.

Figure 5.27: Operator during an opposite side with AR visor test.

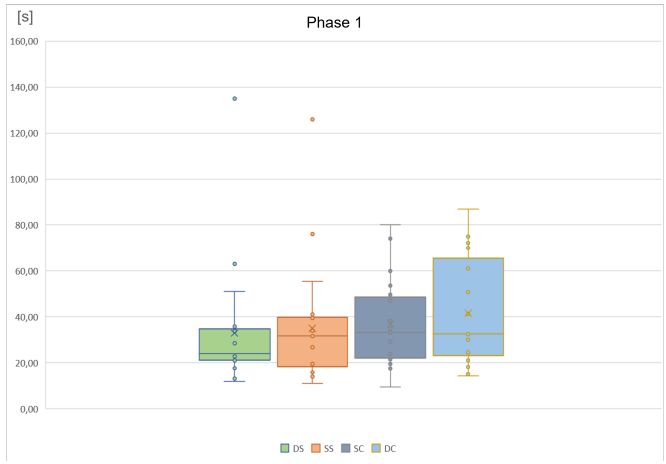


Figure 5.28: Phase 1 (multi-operator) total procedure time by configuration. Total time is defined as detection + extraction per attempt. Configurations: machine-side/monitor, machine-side/AR, opposite-side/monitor, opposite-side/AR.

Despite that, nearly half of the operators valued the ability to freely place the ultrasound image within the workspace as an ergonomic improvement, facilitating continuous attention on the operative field instead of alternating gaze to a fixed screen. Additionally, all participants expressed willingness to recommend the AR approach for further evaluation in practice.

Discussion. The AR monitor behaved as a task-centric, relocatable display that preserves line-of-sight continuity and hand–eye coordination. Qualitative feedback highlighted ergonomic benefits and sustained focus on the puncture field, while also noting areas for refinement (perceived sharpness, initial familiarisation) that can be addressed via display calibration and brief onboarding. Notably, despite all operators being AR-beginners, questionnaire ratings for perceived usefulness and image quality were favourable; this suggests that a short, structured training phase—covering panel-placement heuristics, gesture shortcuts, and brightness/contrast presets—could compress the learning curve and reduce the observed timing overheads, bringing AR-assisted performance closer to conventional monitor-based workflows. Taken together, findings support the feasibility of HoloLens-delivered ul-

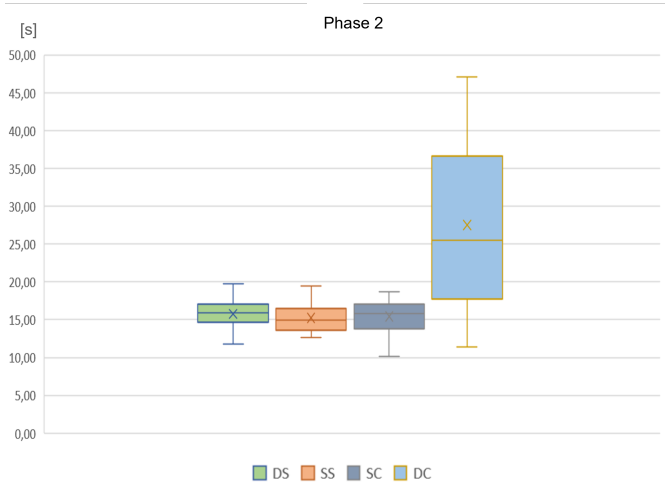


Figure 5.29: Phase 2 (single-operator) total procedure time by configuration. Same definition and ordering as Figure 5.28.

trasound for breast biopsy tasks on benchtop models, without measurable penalties in speed or success.

Limitations and future work. This evaluation used phantoms rather than patients and was not powered for definitive time–efficiency noninferiority; image quality and legibility under varied ambient lighting warrant optimisation. Future steps include larger operator cohorts, clinical validation in real cases, annotation/recording features for collaboration, and expanded training use where AR can shorten the learning curve for ultrasound-guided interventions.

5.4.3 Ultrasound-guided cannulation

This study investigates whether presenting the live ultrasound (US) image *in situ* via a head-mounted display can improve the ergonomics and execution of ultrasound-guided vascular access. The central hypothesis is that anchoring the US view within the operator's field of regard—rather than on a remote monitor—reduces head–eye switching, preserves a continuous line of sight to the needle–target trajectory, and supports stable, sterile workflow.

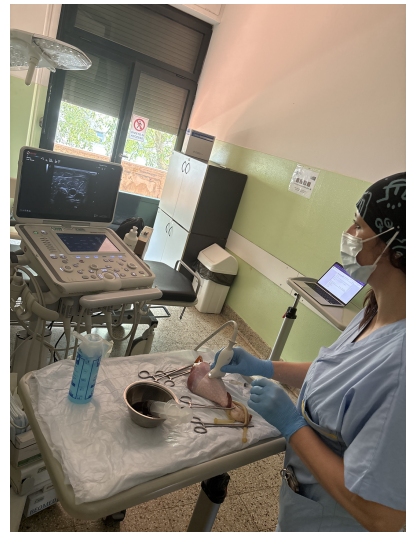
AS for the previously described project, a custom software pipeline in Python was developed. The server captured the video output of a portable ultrasound system, performed lightweight frame handling with OpenCV, and transmitted the stream over a low-latency WebSocket connection. On the client side, a Microsoft HoloLens 2 application decoded and rendered the feed as a spatially anchored, movable panel. Gesture-based controls (with optional voice commands) enabled hands-free repositioning, scaling, and visibility toggling, thereby maintaining sterility.

Face-validity testing was conducted on turkey thighs used as anatomical phantoms, selected for their biomimetic echogenicity. Vascular targets were simulated by tunnelling flexible rubber cannulas longitudinally within the tissue and filling them *in situ* with a povidone–iodine (Betadine) solution, thereby generating a fluid-filled lumen. On B-mode ultrasound, this configuration produced a thin echogenic wall with an anechoic lumen, yielding a clear needle target. A single anaesthesiologist with extensive ultrasound and AR-beginner experience performed all cannulations across two sessions at Careggi University Hospital (Florence, Italy). Each session comprised attempts in both modalities: the traditional monitor-guided workflow and the AR-assisted workflow with the HoloLens 2 device. Illustrative photographs from the set-up and trials are shown in Figure 5.30.

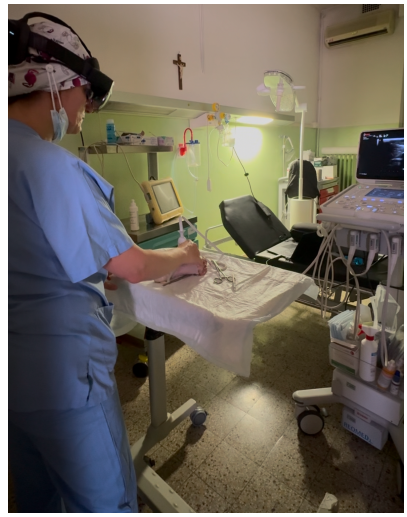
During the experiment, different metrics were measured, including total procedure time, time-to-target, number of needle passes, and qualitative user feedback via a structured questionnaire. Total procedure times for the traditional and AR conditions are reported in Figures 5.31(a)–5.31(b); while a comparative distributional summary is provided in Figure 5.31(c). To explore potential learning effects, the coefficient of determination (R^2) was computed between attempt index and total time within each modality.

Results. Regression analyses yielded consistently low R^2 values in both conditions—R^2=0.0653 (traditional) and R^2=0.0730 (AR)—indicating weak evidence of a learning curve within the small number of repetitions. Nevertheless, summary distributions suggested a modest stabilisation with AR: boxplots showed slightly reduced variability and tighter interquartile ranges for the AR modality. While median total times were broadly comparable between modalities, the failure rate was lower with AR. Qualitatively, this pattern is consistent with improved hand–eye coordination and uninterrupted attention to the puncture field when the US view is presented head-up in

(a) Turkey thigh phantoms prepared for the experiment.

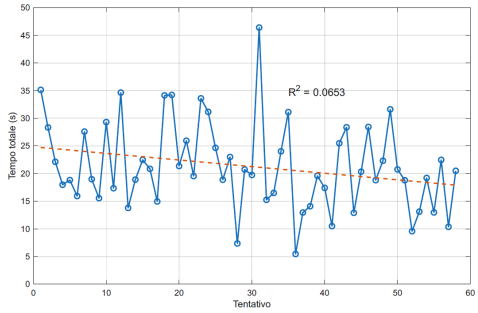


(b) Traditional (monitor-guided) condition.

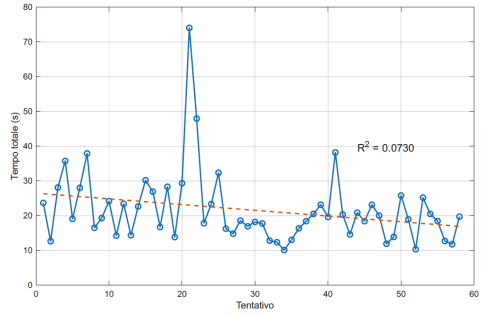


(c) AR-assisted condition with the holographic US panel.

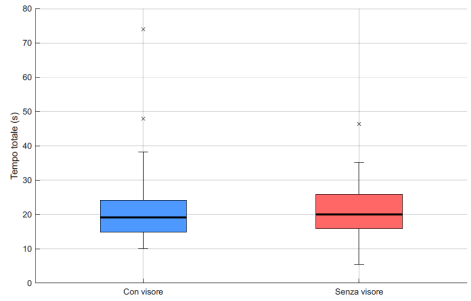
Figure 5.30: Set-up and representative frames from the ultrasound-guided cannulation experiments.



(a) Traditional modality: total time per iteration.



(b) AR modality: total time per iteration.



(c) Box-and-whisker plots comparing total times: AR (left) vs. traditional (right).

Figure 5.31: Timing results for ultrasound-guided cannulation across modalities.

AR. Additionally, the structured questionnaire indicated low perceived difficulty (3/10) and high perceived reliability (9/10) for the AR workflow. The operator reported improved comfort, better focus retention, and enhanced procedural continuity, attributing these to the ability to keep the US image within the natural line of sight and to reposition it on demand.

Discussion. Presenting the US image as a relocatable, head-up display preserved line-of-sight continuity to the puncture field and mitigated gaze toggling between patient and monitor. While AR did not produce a marked acceleration in task execution, the combination of comparable medians, lower failure rate, and narrowed variability aligns with the hypothesis that in-situ visualisation promotes steadier, more consistent performance—an attribute prized in vascular access. Qualitative feedback points to clear ergonomic benefits and sustained attentional focus; incremental gains may be realised with small optimisations to image clarity and initial onboarding.

Limitations and future work. The evaluation used a single expert operator who was AR-beginner and turkey-thigh phantoms rather than patients; statistical power to detect subtle timing differences or learning trends was consequently limited. To minimise bias attributable to lack of prior AR experience, future studies should incorporate a structured familiarisation phase with predefined competency thresholds (e.g., stable panel placement, gesture proficiency) before timed trials, and adopt counterbalanced/crossover designs with a wash-in period to control for learning effects. Larger, multi-operator cohorts spanning experience levels, inclusion of clinical cases, and objective accuracy metrics (e.g., needle-trajectory deviation, first-pass success) are also warranted. Technical refinements—higher effective resolution, further latency reduction, and multi-user visualisation—are planned, together with standardised training to harmonise panel-placement heuristics and gesture usage across operators.

5.5 Organs and tissues segmentation

Accurate volumetric segmentation of anatomical structures from medical images is a foundational task for computer-assisted diagnosis, surgical planning, and quantitative biomedical research. It entails delineating organ boundaries and pathological regions within three-dimensional modalities such as com-

puted tomography (CT), magnetic resonance imaging (MRI), ultrasound, and positron emission tomography (PET). The overarching goal is to transform raw voxels into structured, analysable representations that support both qualitative inspection and quantitative measurement.

Historically, segmentation has been challenging due to anatomical complexity, inter-patient variability, acquisition artefacts, and the often subtle contrast between adjacent tissues. These difficulties are exacerbated in the presence of disease, where standard anatomical assumptions may be inadequate. Consequently, robust, generalisable, and accurate methods have been a long-standing objective in medical image analysis. Classical image-processing pipelines leverage low-level cues and heuristic rules, offering interpretability and modest computational demands but limited capacity to model the full variability of clinical data. In contrast, modern data-driven methods, such as deep neural networks, learn hierarchical features directly from examples and have redefined the state of the art in accuracy and robustness across diverse indications. Fully convolutional architectures, such as U-Net, introduced an encoder–decoder design with skip connections that preserve fine spatial detail during upsampling. Numerous variants (e.g., Attention U-Net, residual U-Net, DeepLab-derived models) incorporate attention mechanisms, residual pathways, and multi-scale context to improve boundary fidelity and robustness. For volumetric data, 3D extensions (e.g., 3D U-Net) exploit inter-slice continuity to enhance anatomical coherence. Additionally, training strategy is as important as architecture: class imbalance—ubiquitous in medical segmentation—motivates loss functions such as (generalised) Dice, focal, or compound losses that emphasise minority structures and boundary accuracy. Data augmentation (rigid and elastic transforms, intensity perturbations, noise models) improves generalisation, while transfer learning and semi-/self-supervised schemes mitigate limited annotation regimes. Rigorous validation typically reports Dice similarity coefficient (DSC), Jaccard index, and boundary metrics (e.g., Hausdorff distance), increasingly complemented by uncertainty estimation and explainability tools to aid clinical adoption.

In the next three sections, we apply high-performance neural segmentation to hepatobiliary–pancreatic surgery, extracting *liver parenchyma*, *hepatic vessels*, and *hepatic tumours* from CT data with the explicit goal of augmented-reality visualisation. Because the liver is densely vascularised, safe resection planning depends on understanding the spatial interplay be-

tween lesions and vascular trees (portal and hepatic venous systems). Rendering co-registered parenchyma–vessel–lesion models as manipulable holograms can sharpen preoperative spatial reasoning—helping surgeons anticipate risk to critical vessels, evaluate margins, and communicate strategy—thereby complementing conventional monitor-based review with an immersive, anatomically faithful 3D context.

5.5.1 Liver parenchyma segmentation

This work sits at the intersection of deep-learning organ segmentation and immersive visualisation, targeting pre-operative planning in augmented reality (AR). Building on the AR platform introduced earlier, we integrate neural network–based segmentation of CT volumes to deliver an end-to-end pipeline for automatic extraction, 3D reconstruction, and *in situ* rendering of organs and tissues on a head-mounted display.

The system adopts a dual-software architecture connecting *3D Slicer* (image computing) and *Unity* (AR visualisation) via the OpenIGTLink protocol. Unity, running on a PC and streaming images to Microsoft HoloLens 2, handles rendering and user interaction; 3D Slicer performs pre-processing and segmentation. A custom Python module in 3D Slicer receives segmentation requests from Unity, executes them, and returns results to the AR client.

At the core of the pipeline is *TotalSegmentator*, a state-of-the-art convolutional neural network derived from the U-Net family. Trained on large, heterogeneous datasets, it predicts masks for more than one hundred anatomical structures (organs, vessels, bones) with high robustness. In our workflow, a CT volume and a user-specified list of target structures are provided to TotalSegmentator (via JSON); the resulting masks are converted to label maps and isolated volumetric datasets for individual organs or organ groups. An example segmentation in 3D Slicer is shown in Figure 5.32.

A key contribution is a fully automated path from segmentation to AR. After inference in 3D Slicer, the volumes are encoded and streamed to Unity, where they are reconstructed and rendered interactively on HoloLens 2. A representative liver segmentation is shown in Figure 5.33.

Given the clinical relevance of hepatobiliary surgery, particular emphasis was placed on the abdomen—specifically the liver and its vasculature. In collaboration with hepatobiliary surgeons (AOU Careggi Hospital), we conducted a case study to assess AR visualisation for hepatic planning. TotalSegmentator was used to segment the liver, portal vein, and inferior

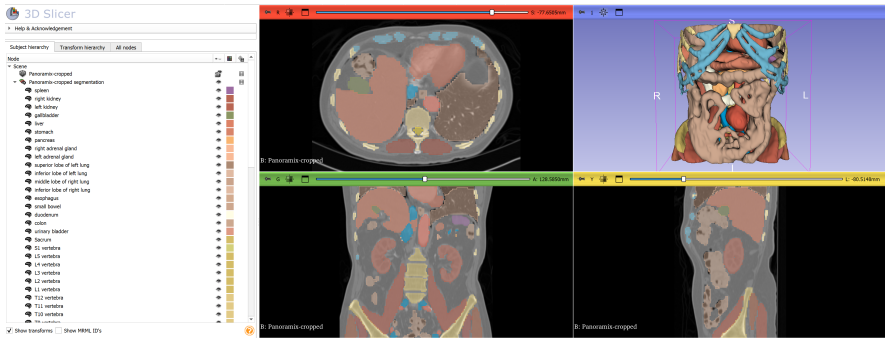


Figure 5.32: Example output of TotalSegmentator visualised in 3D Slicer.

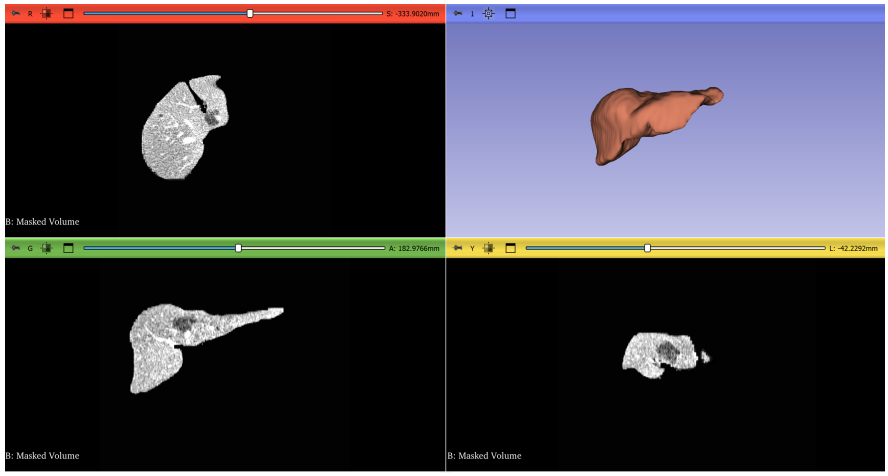


Figure 5.33: Liver parenchyma segmentation obtained with TotalSegmentator.

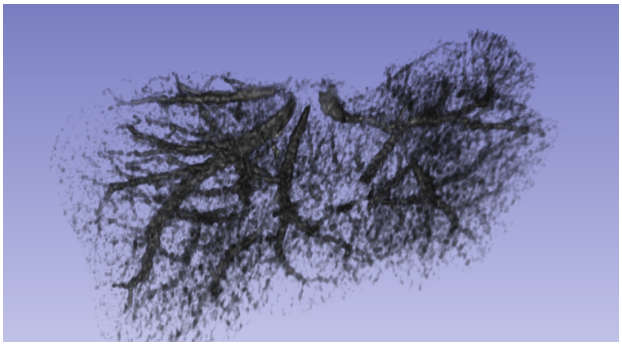


Figure 5.34: Noisy vessel extraction obtained with a threshold-based method (abandoned in favour of transfer-function manipulation).

vena cava. To enhance the depiction of smaller intrahepatic vessels, an initial threshold-based approach was tested (Figure 5.34) but proved sensitive to contrast and noise; we therefore adopted transfer-function manipulation—mapping intensity ranges to colour and opacity—to visually isolate vascular structures from parenchyma.

To support simultaneous inspection of multiple components, the AR application assigns independent transfer functions per segmented object (e.g., parenchyma and vessels) and applies them automatically through a dedicated command in the AR interface. All reconstructions preserve the original CT geometry, ensuring consistent co-registration. An example with liver parenchyma and vessels, inferior vena cava (blue), and portal vein (green) is shown in Figure 5.35.

Each organ can be manipulated independently or as a group using hand gestures—rotate, scale, and dissect—directly within the user's field of view. Figure 5.36 illustrates typical interactions: multi-object selection with locking, global and selective clipping, and lesion visualisation.

In sum, the parenchyma module was the first step of the pipeline. While liver parenchyma segmentation—and the large-vessel masks for the inferior vena cava and portal vein—were consistently reliable, intraparenchymal tumours were not segmented and were thus subsumed within the liver label. Likewise, our initial attempt to approximate smaller hepatic vessels via ad-hoc transfer-function tuning yielded a visually suggestive yet clinically inadequate result. These limitations motivated two dedicated extensions: (i) a vessel-centric workflow to segment and render the hepatic vascular trees,

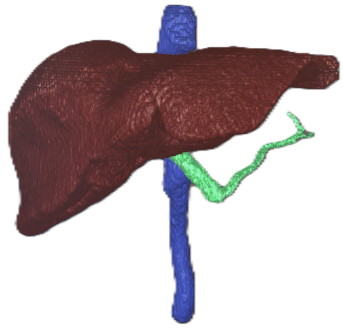


Figure 5.35: Co-registered reconstructions in AR: liver (parenchyma and vessels), inferior vena cava (blue), and portal vein (green).

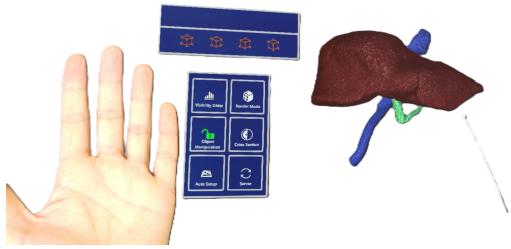
and (ii) a lesion-centric workflow to detect and isolate hepatic masses. The following sections detail these additions.

5.5.2 Liver vessel segmentation

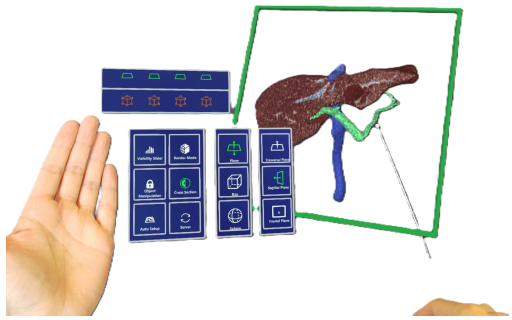
To address the limitations of parenchyma–only visualisation, we developed a dedicated pipeline for hepatic vessels segmentation and holographic display. The solution couples automated, data–driven extraction of the portal/hepatic venous trees from contrast–enhanced CT with our Unity–3D Slicer framework, enabling in-situ inspection on HoloLens 2 and coordinated interaction with liver parenchyma and (in later stages) tumour segmentations.

As in the preceding modules, 3D Slicer acts as the processing back end and Unity as the AR client. Segmentation requests are issued from Unity, executed in 3D Slicer, and returned via OpenIGTLink as labelled volumes or meshes ready for rendering. Once received, vessels can be visualised alone or co-rendered with other structures, with linked manipulation and shared clipping planes to interrogate spatial relationships.

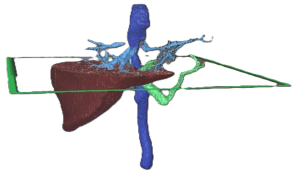
A public dataset, called Liver Tumour Segmentation (LiTS) Challenge (3D CT with expert labels for background, liver, and tumour), was used as training data for the neural network and comprised contrast CT volumes with paired vessel annotations (NIfTI format). All studies were reviewed in 3D Slicer to verify image–label alignment and basic integrity. To stabilise learning for a topology-sensitive target, we framed the task as binary vessel extraction: any ancillary labels (e.g., tumours) were remapped to



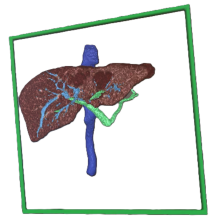
(a) Hand menu with all volumes selected (red boxes) and manipulation enabled (green padlock).



(b) Global clipping plane applied to all volumes.



(c) Selective clipping of liver parenchyma only.



(d) Multiple hepatic lesions (darker regions).

Figure 5.36: Representative manipulations and interactions with segmented volumes in AR.

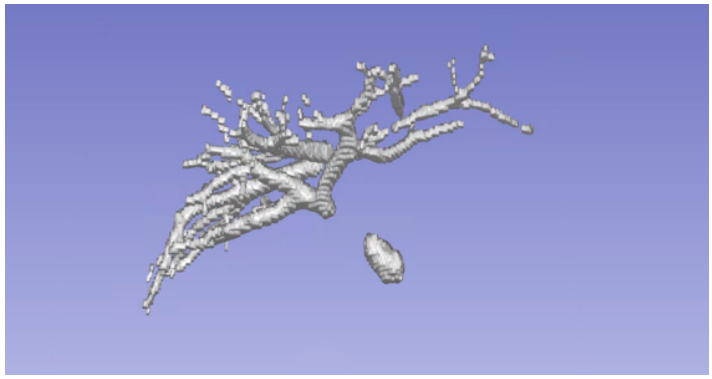


Figure 5.37: Liver vessels segmentation result obtained using the trained nnU-Net v2.

background. A .json file described modality, channel mapping, and class definitions according to nnU-Net conventions, and the canonical directory structure was created to support reproducible experiments.

We adopted nnU-Net v2 in its 3D full-resolution configuration to preserve small calibres and tortuous geometry. Following automatic fingerprinting and plan generation, the network was trained end-to-end on the curated cohort with extensive on-the-fly augmentation (affine, elastic, and intensity perturbations) to improve robustness to protocol variability. Dice-oriented compound losses were used to counter severe class imbalance; post-processing removed isolated false positives and enforced connectivity where appropriate. On held-out cases, the model consistently recovered the portal trunk and first–/second–order branches with anatomically plausible branching and continuity. As expected, recall declined in distal, low-contrast segments, where occasional fragmentation and small gaps appeared—reflecting voxel resolution, contrast heterogeneity, and extreme foreground sparsity. After isosurface extraction, the resulting meshes proved suitable for anatomical review and downstream AR exploration. Liver vessels segmentation result, shown in 3D Slicer, is shown in Figure 5.37.

Segmented vessel volumes/meshes are streamed to Unity and instantiated alongside the liver parenchyma. Users can toggle structures, adjust independent transfer functions, and apply synchronised clipping planes across all objects to expose vessel–parenchyma interfaces and candidate resection planes. Standard MRTK gestures (grab, rotate, scale) support single- or

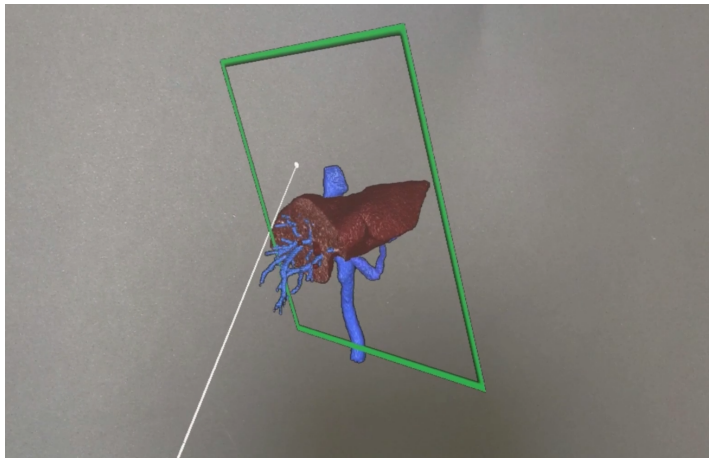


Figure 5.38: User slicing segmented volumes in AR.

multi-object manipulation, with anchored layouts for team discussion or pre-operative briefing. An example of a user interacting with segmented volumes is shown in Figure 5.38.

With this second step, we moved beyond parenchyma–only visualisation: liver parenchyma and the hepatic venous trees are now produced as separate, co-registered volumes and rendered as independent holograms. In AR, this enables explicit interrogation of their spatial interplay—toggling visibility, applying synchronised clipping planes, and inspecting candidate resection corridors with vessels in situ. To complete the clinical triad required for hepatobiliary decision-making, the next module targets intra-parenchymal tumour segmentation, so that lesions, vessels, and parenchyma can be analysed together within a unified AR workspace.

5.5.3 Liver tumour segmentation

This last step of the segmentation project tackles the clinically critical task of segmenting hepatic tumours from computed tomography (CT) using deep learning. We adopted nnU-Net v2, a self-configuring framework that adapts preprocessing, architecture, and training to the data at hand, and trained it on the Liver Tumour Segmentation (LiTS) Challenge dataset (3D CT with expert labels for background, liver, and tumour).

All CT volumes and labels were validated for geometric consistency and

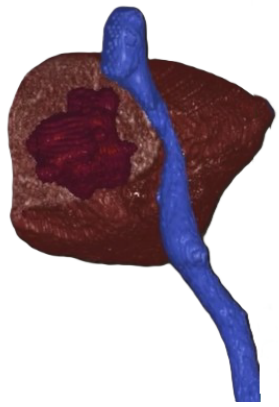


Figure 5.39: Visualisation of liver (soft red), tumour (hard red) and vena cava (light purple) segmentations in the AR environment.

reorganised to comply with nnU-Net's canonical structure. A .json metadata file specified modality, class mapping, and channel layout. Training used the 3D full-resolution configuration on a CUDA-enabled workstation (Windows 10, Python 3.10, PyTorch/CUDA 12.1), with 5-fold cross-validation (1,000 epochs per fold). The framework automatically produced best-checkpoint and final models per fold according to Dice performance.

Liver masks were recovered robustly, with Dice similarity coefficients (DSC) typically in the 0.71–0.79 range across folds. Tumour segmentation—intrinsically harder due to small size, irregular morphology, and low contrast—achieved DSC between 0.38 and 0.57. The best-performing fold (fold 3) balanced liver and tumour accuracy and was selected for downstream use.

Inference was first executed via the nnU-Net CLI and subsequently through a dedicated nnU-Net extension in3D Slicer (v5.8.1), enabling GUI-based prediction and immediate inspection in the Segment Editor. Segmented volumes were then integrated into our AR pipeline: 3D Slicer and Unity communicated via OpenIGTLink through Python, which serialised segmentation requests and streamed labelled volumes for holographic rendering on HoloLens 2. Examples of interactive AR inspection—including clipping through liver, tumour, inferior vena cava, and hidden intrahepatic vessels—are shown in Figure 5.39.

This module completes the hepatobiliary triad introduced above: first step delivered accurate parenchyma (and major veins), second one added dedicated vessel extraction, and the present step contributes lesion masks. Together, parenchyma, vessels, and tumours are produced as separate, co-registered volumes and rendered as manipulable holograms, enabling surgeons to interrogate vessel–lesion–parenchyma relationships in situ—toggling structures, synchronising clipping planes, and exploring candidate resection corridors—within a unified AR workspace for pre-operative spatial reasoning, education, and multidisciplinary discussion. Demonstrations to our collaborators at Careggi University Hospital were met with strong enthusiasm: clinicians valued the ability to co-inspect vasculature and lesions in true 3D and saw clear potential for briefing and teaching. More broadly, while our deployment focused on hepatobiliary–pancreatic surgery, the same pipeline naturally extends to other domains (e.g., neuro-oncology: tumour with eloquent tracts and vessels; orthopaedics: bone with neurovascular bundles; cardiac: myocardium with coronaries; head-and-neck; paediatrics), simply by swapping the trained segmenters and transfer functions.

At present, the system supports pre-operative planning with precomputed segmentations and interactive AR exploration. Next steps include (i) quantitative validation in prospective studies, (ii) improved small-lesion sensitivity via hybrid losses and multi-phase CT fusion, (iii) semantic vessel labelling (portal vs hepatic venous trees, segmental branches) with volumetry and margin tools, (iv) patient-space registration for intraoperative use (fiducials/surface ICP; ultrasound/fluoro co-registration), and (v) multi-user session synchronisation and ergonomic refinements. Together, these enhancements aim to translate AR-assisted, segmentation-driven visualisation from a compelling prototype to a robust clinical decision aid.

5.6 Visualisation of 4D medical volumes

This section describes the design and implementation of an augmented reality (AR) system for the interactive visualisation of 4D medical volumes (3D + time). Whereas earlier chapters addressed only static volumetric datasets, the present work extends the pipeline to support dynamic 4D datasets within the same AR environment.

The implementation followed a two–stage development strategy:

- **Server–side tooling:** was done in 3D Slicer for loading, formatting,

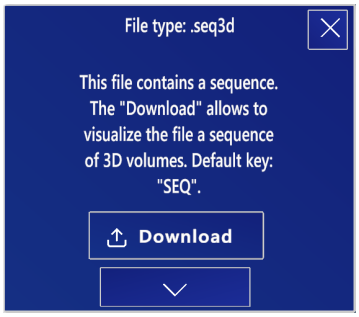


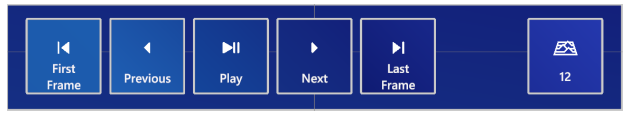
Figure 5.40: User hand menu when a `.seq3d` file is selected for download from the server.

and serving volume sequences;

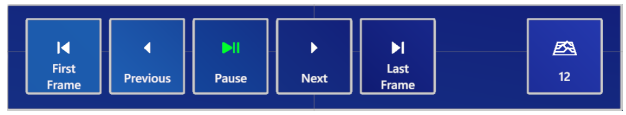
- **Client–side extensions:** was done in Unity for real–time presentation on HoloLens 2 via Holographic Remoting.

Together, these components form a complete pathway for importing, serialising, transmitting, and interactively rendering time–dependent 3D medical data in AR.

The system adopts a client–server paradigm. 3D Slicer, running on a remote Linux server, handles sequence ingestion, preprocessing, and serialisation. Unity, remoted to HoloLens 2, acts as the AR client, receiving formatted data and rendering it for in–situ exploration. A key enhancement is a custom container format, `.seq3d`, designed for structured serialisation of 4D datasets (temporally ordered 3D frames). Each file begins with a 1024 byte header that encodes the frame count, data type, and byte offsets, enabling consistent, efficient deserialisation and frame reconstruction on the client. A Python module was developed in 3D Slicer to (i) import sequence files; (ii) split them into per–timepoint volume nodes; and (iii) encode the entire series into a single `.seq3d` archive. The module also supports the inverse operation, converting a stored `.seq3d` into a Slicer scene. Server–client communication is established via OpenIGTLink: before transmission, a text node is sent to Unity to advertise an incoming, structured multi–frame payload, allowing the client to allocate resources appropriately and preserve data integrity during streaming. An AR hand menu for selecting and downloading a `.seq3d` file from the server is shown in Figure 5.40.



(a) Playback menu while volume rendering is paused.



(b) Playback menu while the sequence is running.

Figure 5.41

The Unity client was extended to recognise time–based volumetric payloads alongside static volumes. Based on the metadata carried by the initial text node, the client routes the stream to the appropriate object class. Rendering combines MRTK interaction components with custom shader–based volume ray–casting, preserving the previously available tools (3D manipulation, cross–sectioning) and adding temporal navigation. A dedicated playback interface (Figure 5.41) supports framewise stepping (first/previous/next/last) and continuous playback at configurable frame rates (12, 24, 30, 60 FPS). Users can translate, scale, and rotate volumes and place clipping planes in real time, with the controls rendered directly in AR space.

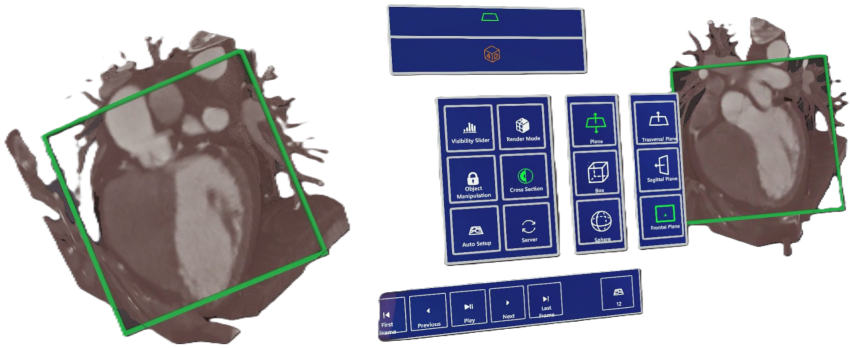
Once received, sequence data are cached on–device and synchronised with user inputs, enabling smooth, immersive interaction even for temporally dense datasets (Figure 5.42).

For functional validation, a dynamic cardiac CT sequence was used. The system successfully loaded, played back, and manipulated 4D datasets with low perceived latency and stable image quality, while preserving the full interaction toolkit (placement, scaling, rotation, and slicing).

At present, serialisation stages create temporary per–frame files, which is suboptimal for very long sequences. In addition, generating a `.seq3d` file and uploading it to the server are separate user actions. Future improvements will target: (i) memory–efficient, chunked streaming with on–the–fly decompression; (ii) unified one–click packaging and deployment; (iii) adaptive quality control (dynamic resolution/frame–rate scaling); and (iv) optional remote rendering or multi–user synchronisation for collaborative review.

(a) Volume rendering in the AR environment.

(b) Same volume rendering with the user hand menu enabled ("object manipulation" active, green icon).



(c) Volume rendering cut using a clipping plane in the AR environment.

(d) Same volume with the "frontal cross–section plane" option enabled (green icons).

Figure 5.42

5.7 Web-based 3D mesh viewer with real-time AR co-manipulation

Augmented reality (AR) headsets typically provide a *single-user* view: the wearer sees and manipulates the hologram, while colleagues without a headset cannot share the same perspective in real time. This limits team situational awareness, teaching, and quality assurance, and it complicates telemedicine scenarios where remote participants need to observe, annotate, or steer the interaction. Hybrid collaboration—linking head-mounted AR with ubiquitous, browser-based viewers—offers a pragmatic remedy: state (mesh transforms, slicing planes, annotations, points of contact) can be mirrored across modalities so that on-site and remote users co-exist in one shared workspace.

Motivated by these constraints and opportunities, during a three-month research visit to the *Centre for Digital Medicine and Robotics*, Jagiellonian University Medical College (Krakow, Poland), we designed a multi-component system for real-time, cross-platform visualisation and co-manipulation of 3D meshes spanning the web and AR. The solution comprises three coordinated parts—server, web client, and AR client—engineered for low-latency synchronisation and collaborative use. A functional schematic is shown in Figure 5.43.

A lightweight server was implemented in C++ (Qt), exposing bidirectional WebSocket channels to heterogeneous clients. The server broadcasts state updates (e.g., mesh load, transform, clipping parameters) and relays interaction events, enabling synchronous multi-endpoint visualisation with minimal latency. Then, a browser application was developed in Qt and compiled to WebAssembly for broad, installation-free access. Users can load arbitrary 3D meshes and interact via standard mouse navigation (orbit/pan/-zoom). Rendering is performed with OpenGL, and an adjustable cutting plane permits arbitrary slicing (normal, offset, and on/off control) directly in the viewport. Integrated WebSocket modules keep the web view synchronised with remote AR interactions and server events. Examples of the interface and slicing tool are shown in Figure 5.44.

Finally, an AR application was built in Unity (C#) for HoloLens 2. Through WebSocket connectivity, the client requests the same mesh loaded in the web app and renders it as a hologram. Users manipulate the model with MRTK gestures (grab, rotate, scale) and apply an AR cutting plane

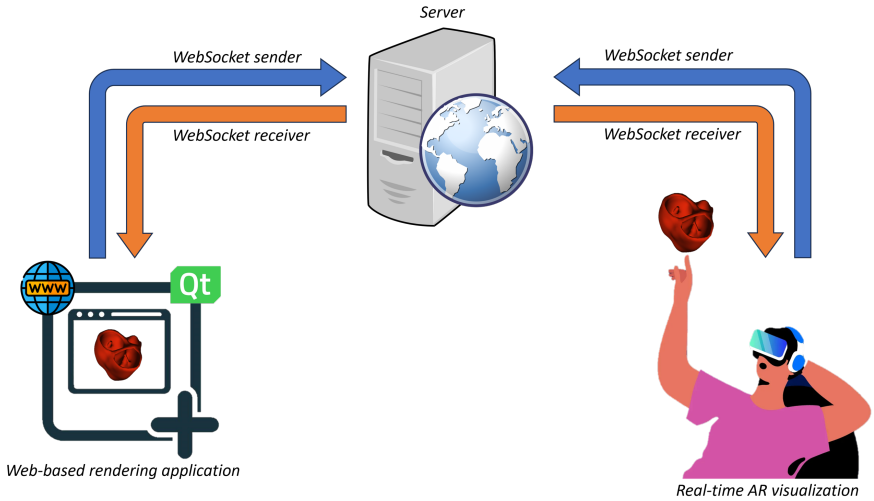


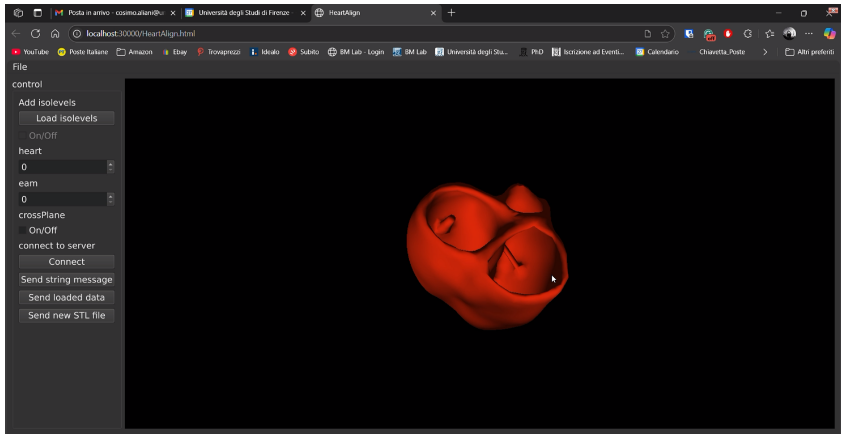
Figure 5.43: System overview: a C++/Qt server (WebSockets) brokers low-latency communication between a browser-based WebAssembly client (OpenGL renderer) and a HoloLens 2 AR client (Unity/MRTK).

analogous to the web tool. For collaborative awareness, hand raycast hit points detected in AR are transmitted to the server and visualised in the web client, creating a shared interaction space across modalities. Representative interactions are shown in Figure 5.45.

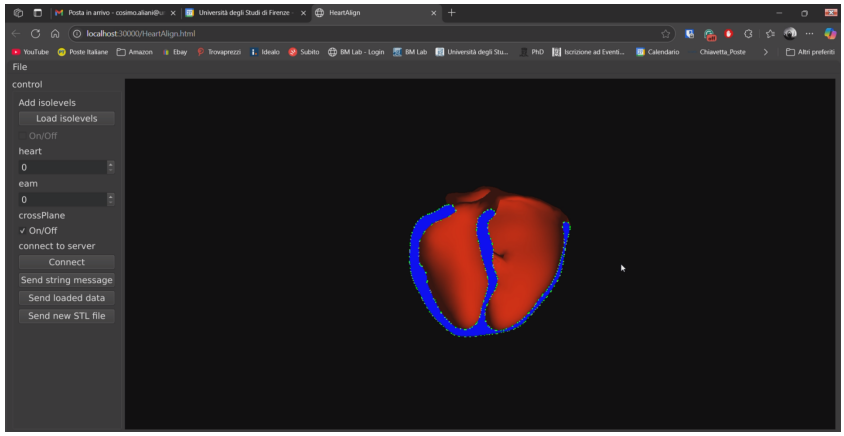
The resulting platform bridges web-based rendering and immersive AR manipulation over a common WebSocket backbone, supporting synchronous state sharing and reciprocal annotation. Beyond consolidating expertise in C++, C#, Qt, WebAssembly, OpenGL, and Unity/MRTK, the project informs the design of multi-modal, collaborative visualisation systems and suggests clear future extensions (e.g., multi-user state replication, persistent sessions, compression for large meshes, and role-based access) for clinical and educational deployment.

Chapter synthesis and concluding remarks

This chapter can be synthesised as the gradual construction of an AR medical workspace: a set of interoperable components that allow clinicians (and

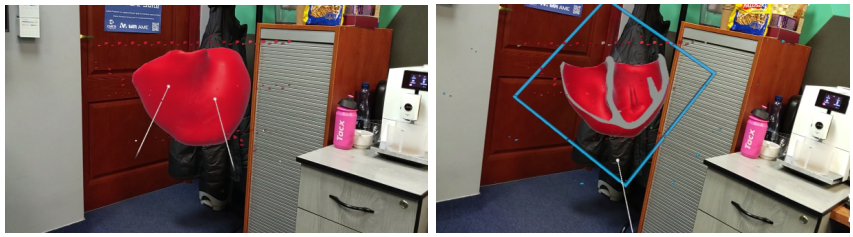


(a)

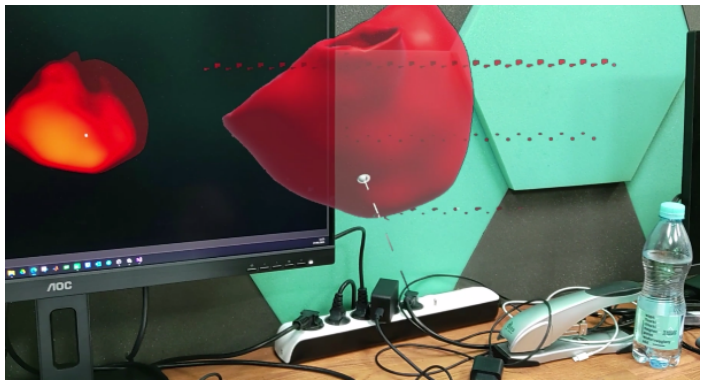


(b)

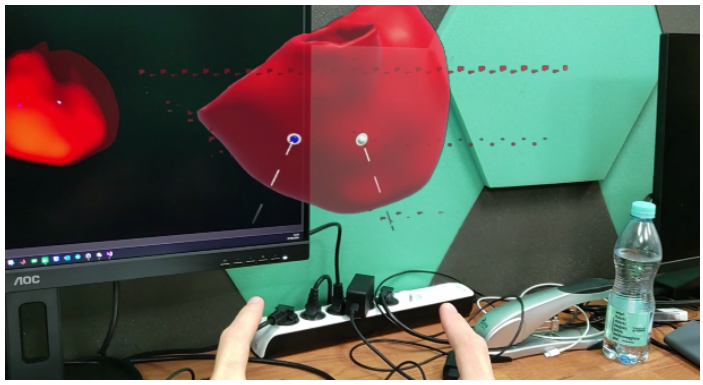
Figure 5.44: Web client (WebAssembly/OpenGL): (a) interactive mesh inspection with mouse navigation; (b) configurable cutting plane for arbitrary cross-sections.



(a) (b)



(c)



(d)

Figure 5.45: Human–hologram interaction on HoloLens 2: (a) two-hand manipulation; (b) AR cutting plane; (c–d) one- and two-hand raycast hit points mirrored to the web client for shared context.

technical staff) to access, understand, and act on 3D/4D medical information directly within the spatial context of care.

The first contribution is a robust volumetric rendering and interaction core on HoloLens 2. Beyond basic holographic display, the system supports multiple rendering modes, transfer-function control, and cross-sectional interrogation, establishing the practical primitives required to transform clinical volumes into manipulable spatial artefacts. The subsequent extensions, remote dataset browsing, asynchronous download, and patient-list status management, shift the system from a standalone demonstrator to a workflow-aware tool: AR becomes capable of operating on organised, patient-indexed data rather than isolated files. Second, the QR-code registration module addresses the non-negotiable requirement that AR must be spatially repeatable when the intent moves toward guidance or education with physical phantoms. The value of this contribution lies not only in the chosen tracking strategy, but in demonstrating a pragmatic pathway that trades external infrastructure for a headset-native inside-out solution while retaining usability and stability in practice. Third, the digital palpation project extends the interaction vocabulary beyond rigid transforms, introducing a deformation model that supports multi-finger and bimanual manipulation under real-time constraints. In parallel, the ultrasound projects reframe AR from "3D reconstruction" to "human-factors augmentation": the live ultrasound feed is relocated into the operator's field of view as a spatially anchored, repositionable panel. Even when speed gains are not dominant, reduced attention switching and improved procedural continuity are clinically meaningful outcomes, especially in time-pressured or ergonomically constrained settings. Fourth, the segmentation pipeline completes a clinically motivated triad – parenchyma, vessels, tumours – and connects it to AR through a modular bridge between 3D Slicer (processing) and Unity/HoloLens 2 (visualisation and interaction). Conceptually, this shifts the AR environment from displaying pre-prepared models to supporting an end-to-end path where patient-specific anatomy can be generated, updated, and explored with minimal friction. Finally, the 4D volume framework expands representational capacity to time-resolved datasets, while the web–AR co-manipulation platform expands the collaboration model: AR no longer needs to be an isolated, headset-only experience, but can be mirrored to ubiquitous web clients so that non-headset users (local or remote) can observe, annotate, and follow the same interaction state.

In aggregate, the projects in this chapter advance AR from a sequence of

isolated prototypes to a coherent technical direction: a reusable, extensible AR layer that supports medical visualisation, interaction, alignment, workflow connectivity, and collaborative review. This provides the conceptual and engineering foundation needed to consider AR not as a "demo modality", but as an enabling interface for future clinically validated tools.

Part III

Discussion and conclusions

Chapter 6

Discussion

A superficial reading of Chapters 4 and 5 might suggest a collection of heterogeneous prototypes. However, the projects can be interpreted as parts of a coherent research direction centred on a practical problem: *how to make medical data usable in physical space under realistic clinical constraints.* Across both optical 3D imaging and head-mounted augmented reality (AR), clinical usefulness depends less on compelling demonstrations and more on geometric trust: the ability to establish and preserve known reference frames, to quantify alignment quality, and to understand how performance degrades under adverse conditions.

Within this perspective, optical 3D imaging and AR are not necessarily "parallel topics" but can be seen as complementary layers. Optical sensing provides tools to capture, register, and validate geometry in the real world; AR provides an interaction and visualisation layer that can externalise complex 3D information into the clinician's workspace. The overall contribution can therefore be described as a progressive construction of reusable components—sensor selection rules, fiducial-based registration strategies, streaming/remoting pipelines, and modular visualisation modules—and their evaluation in clinically motivated scenarios.

6.1 Recurring considerations across the projects

6.1.1 Robustness in clinical settings often matters more than best-case accuracy

A recurrent observation is that high nominal accuracy is of limited value if the sensing modality fails in common clinical conditions. Depth cameras do not interact with "generic objects" but with metallic tools, retroreflective markers, drapes, and low-texture surfaces. Under these conditions, predictable degradation is often more valuable than occasional millimetric precision. This perspective supports modality choices that prioritise stable operation over best-case performance, and it motivates explicit documentation of failure modes as an important output rather than a minor detail.

6.1.2 Reference frames and registration tend to be the main practical constraint

Many projects converge on a shared constraint: the limiting factor is rarely rendering quality or algorithmic sophistication, but the capacity to define and maintain a repeatable mapping between (i) the physical world, (ii) sensor measurements, and (iii) patient-derived imaging coordinates. Fiducial-based strategies (e.g., ChArUco, QR-based anchors) repeatedly emerge as a reliable route to repeatability, particularly when scenes lack robust texture or include reflective surfaces. The trade-off is operational: fiducials impose line-of-sight and setup requirements, which must be justified by the clinical intent (education versus guidance versus intra-procedural use).

6.1.3 System architecture choices are closely tied to latency and compute constraints

Both AR and optical pipelines are sensitive to latency, but latency is not only a performance metric: it shapes interaction, perceived stability, and ultimately user trust. The work therefore repeatedly adopted architectural strategies that relocate computation to where it is cheaper and more predictable: desktop-hosted rendering via remoting, server-side ultrasound acquisition with ROI control, and back-end segmentation in 3D Slicer with protocol-based transfer to Unity. These choices reflect a view of AR headsets

as interaction devices rather than fully self-contained compute platforms, at least for demanding workloads.

6.1.4 In some workflows, the main benefit is related to human factors

Not all clinically meaningful gains manifest as faster task completion. In procedure-adjacent contexts (e.g., ultrasound-guided interventions), AR's most defensible value proposition may be attentional: reducing head-eye switching, preserving line-of-sight continuity, and lowering cognitive friction. In such cases, comparable procedure times can still be interpreted as a positive outcome if variability and failure rates decrease, or if users report improved comfort and focus. This has implications for evaluation: success should be measured not only in speed, but also in consistency, error patterns, and perceived workload.

6.1.5 Reusing components across tasks appears feasible and useful

Across the dissertation, a small set of building blocks—robust pose estimation, multi-view registration, volumetric rendering with slicing, low-latency streaming, and bidirectional communication between processing back ends and AR clients—was repurposed across different projects. The value of this approach lies in showing that once these components are characterised, they can be adapted to distinct workflows (positioning, texturing, reslicing, planning, collaboration) without redesigning the entire system from scratch.

6.2 Interpretation of results by intended use

6.2.1 Work focused on geometry: positioning and verification

Projects that target positioning and geometric verification can be discussed primarily through the lens of safety and operational efficiency. For example, replacing ionising scout views with non-ionising optical verification is compelling not because it is novel, but because it addresses an everyday clinical friction: repeated positioning checks that cost time and dose. In this category, the most relevant outputs are (i) quantified spatial fidelity, (ii) clearly

described failure conditions, and (iii) feasibility of integration within existing hardware constraints.

6.2.2 Work focused on visual understanding and communication

Texture mapping of radiological models and AR-based volumetric exploration can be discussed as interventions on human understanding. Realistic appearance and manipulable 3D representations can support education, consent, multidisciplinary discussion, and preoperative rehearsal. The appropriate benchmark here is not diagnostic accuracy, but whether the representation improves recognisability, communication efficiency, and shared spatial reasoning. This also motivates web–AR hybrid collaboration as a practical strategy: it reduces the exclusivity of headset-only experiences and can support team-based adoption.

6.2.3 Work adjacent to procedures: ergonomics and consistency

In ultrasound-guided interventions, AR can be discussed as a human-factors layer: the system does not change the ultrasound signal, but changes how the signal is accessed at the point of action. The discussion therefore concerns attentional continuity, ergonomic adaptability to room layout, and the learning curve of gesture-based interaction. Observed timing overheads in AR-beginner cohorts can be interpreted as a training effect rather than a fundamental limitation, provided that variability and failure patterns trend favourably and that clear onboarding strategies exist.

6.2.4 Work supporting planning: segmentation-to-AR pipelines

The segmentation-driven hepatobiliary workflow can be discussed as a step toward planning support: parenchyma, vessels, and tumours become co-registered, manipulable entities that clinicians can interrogate in true 3D. The key question is not whether AR is visually impressive, but whether the pipeline reduces friction from imaging to spatial understanding, supports more explicit reasoning about margins and vascular risk, and improves communication within the surgical team. Importantly, the limitations of tumour

segmentation accuracy and generalisability must be treated as central constraints on the strength of any clinical claims.

6.3 Limitations and practical constraints

Beyond the specific outcomes of the individual projects presented in this dissertation, a number of limitations and areas for improvement remain. These limitations span physical constraints of optical sensing, practical constraints of deployment, and methodological constraints of the current evaluations. Discussing them explicitly is important both to interpret the results appropriately and to outline realistic directions for further development.

Optical depth sensing remains vulnerable to material-dependent failures (specularity, retroreflection, translucency), and these failures are not rare edge cases in clinical environments. AR headsets introduce additional sensing constraints: inside-out tracking can degrade under poor lighting or feature-poor settings, while marker-based anchoring depends on visibility and stable detection. A clinically credible system should therefore communicate confidence and degrade gracefully (e.g., warn when tracking quality falls below thresholds).

Several prototypes rely on external compute (remoting, server back ends) or dedicated setup elements (fiducials, printed mounts, capture devices). These dependencies are not inherently negative, but they impose adoption costs: network reliability, sterilisation-compatible placement, and time-to-setup become decisive factors. Future systems should therefore report these operational costs alongside performance metrics.

A significant portion of the evaluation was performed on phantoms, benchtop models, and small user cohorts. While appropriate for early-stage validation, these studies cannot support strong clinical effectiveness claims. Moreover, AR-beginner participants introduce confounding learning effects, especially when gesture fluency and display calibration influence performance. The implication is not that the approach is weak, but that the next phase should prioritise study designs that better separate training from efficacy (counterbalanced crossover designs, wash-in phases, competency thresholds, and clinically meaningful endpoints such as first-pass success and trajectory deviation).

Segmentation pipelines inherit the limitations of their training data and can fail silently when confronted with out-of-distribution anatomy or acqui-

sition protocols. For translation, the system should incorporate uncertainty-aware outputs, quality control checks, and clear user pathways for correction. Without these, an AR interface risks amplifying misplaced confidence by rendering uncertain structures with high visual authority.

6.4 Considerations for clinical translation

Moving from prototype to clinical tool requires aligning the technology with workflow ownership (who initiates it, when, and why), integrating with hospital infrastructure (PACS, data governance, cybersecurity), and establishing a validation ladder from benchtop realism to prospective clinical studies. In parallel, regulatory readiness requires an early articulation of intended use and risk: systems that influence procedural decisions, patient positioning, or intraoperative guidance face stricter expectations than educational viewers. Consequently, translation should proceed with an explicit risk management strategy, usability engineering in representative settings, and security-by-design when streaming or server-based processing is involved.

6.5 Summary of the discussion

The dissertation can be read as an exploration of how optical 3D imaging and AR can be engineered and evaluated for clinically motivated scenarios. Across the projects, the recurring technical requirement is reliable control of reference frames and a clear understanding of failure modes, while the recurring translational requirement is to manage setup burden, latency, and usability. The individual projects are therefore best interpreted not as isolated applications, but as instances where a set of reusable components was tested across different intended uses—from positioning and verification to planning support and procedure-adjacent visualisation.

Additionally, the work described herein not only advances existing methods but also formulates, validates, and integrates new approaches that have been demonstrated in real-world clinical settings. Importantly, several of these advancements have been published in peer-reviewed journals and conferences, highlighting the scientific contribution of this dissertation to the field. The publications emerging from this work provide an explicit demonstration of how the research addresses critical gaps in clinical practice by making existing technologies operational under clinical constraints. These

papers collectively illustrate how design knowledge was extracted and iteratively refined through a series of projects, each building on the last, to develop clinically feasible, reliable, and scalable workflows.

Chapter 7

Conclusions and future directions

Taken as a whole, this doctoral project is not merely a catalogue of isolated devices or algorithms; it is an argument for *task-centred spatial computing* in medicine. Its animating premise is simple: optical 3D imaging and augmented reality (AR) become clinically meaningful only when deployed *where* they relieve genuine pain points—cognitive, ergonomic, or organisational—and only when their coupling yields capabilities that neither strand delivers alone. By moving beyond the novelty of visualization, the individual projects presented in this thesis collectively advance the field towards a unified framework: they demonstrate that clinical utility arises not from maximizing technological complexity, but from optimizing the fit between digital overlays and physical tasks. Consequently, the work privileges pragmatism over spectacle: reproducible pipelines, accessible hardware, and designs that respect the messy constraints of clinical environments.

Abstracting beyond individual case studies, the thesis distils a compact *pattern language* for spatial computing in medicine:

- **Anchor reality early.** Robust spatial alignment is the currency of every downstream capability; invest in it first.
- **Favour perceptual plausibility before physical fidelity.** In many clinical tasks, a stable and legible hologram outperforms a perfectly simulated one that is slow or fragile.
- **Treat collaboration as a core requirement.** Hybrid (web + AR)

experiences expand participation, reduce single-user bottlenecks, and smooth adoption.

- **Build for auditability and scale.** Portable formats, server–client separation, and standards-oriented interfaces ease integration into existing clinical ecosystems.

7.1 Future perspectives

While this thesis establishes a technical foundation, the path from academic prototype to standard of care requires addressing broader systemic challenges. The future development of this research line should focus on three synergistic pillars: clinical translation, regulatory compliance, and continued technical refinement.

7.1.1 Clinical translation and workflow integration

The next phase must move beyond feasibility studies towards rigorous validation of clinical value and user adoption.

- **Validation in powered studies.** Future trials must pivot from technical metrics (e.g., overlay error) to clinical outcome measures. For AR-assisted ultrasound and biopsy navigation, this implies prospective, randomised studies quantifying time to cannulation, first-pass success rates, and reduction in needle path deviation. Similarly, optical scouting tools must be evaluated against standard-of-care regarding patient positioning success and ionizing dose reduction.

- **Workflow integration and adoption.** A technically sound device will fail if it disrupts the operating room's choreography. Future work must assess the integration of these tools into existing hospital information systems (HIS/PACS) and physical workflows. This includes minimizing setup time (the "time-to-first-image"), automating patient registration without cumbersome markers, and designing structured onboarding programs to mitigate the learning curve effects observed in early trials.

7.1.2 Regulatory and translational barriers

Transitioning these systems from research tools to medical devices introduces significant regulatory and ethical complexities that must be proactively addressed.

- **Medical Device Regulation (MDR).** To achieve CE marking or FDA clearance, the software pipelines developed here (e.g., `.seq3d` streaming, segmentation networks) must transition to strict quality management systems (QMS). This involves rigorous risk management, documentation of software lifecycles (IEC 62304), and clinical evaluation reports ensuring compliance with the EU Medical Device Regulation (MDR 2017/745).
- **AI Regulation and Governance.** As segmentation and registration increasingly rely on deep learning, compliance with emerging frameworks like the EU AI Act becomes critical. Future developments must ensure explainability and robustness, particularly for "high-risk" AI systems involved in surgical guidance.
- **Data Privacy and Security.** The proposed hybrid architectures (cloud/edge processing) require robust data protection strategies. Implementing privacy-preserving streaming, end-to-end encryption, and GDPR-compliant data handling is essential, particularly when transmitting patient-specific geometries or video feeds for remote consultation.

7.1.3 Technical roadmap

Finally, specific technical gaps identified during this work offer fertile ground for engineering improvements:

- **Registration and tracking.** Enhancing patient-space registration via hybrid approaches (fiducials + surface ICP), continuous drift checks, and dynamic compensation for soft-tissue motion (US/fluoro updates).
- **Advanced segmentation.** implementing multi-phase CT fusion, topology-aware losses, and uncertainty-aware visualisation for tumours and distal vessels.

- **Systems performance.** Optimizing edge/cloud co-processing (zero-copy streaming) and adaptive quality control. Hardening the `.seq3d` pathway with chunked streaming and on-the-fly decompression will be key for scalability.
- **Human–Computer Interaction (HCI).** Standardising hand-menu layouts and exploring spatial audio cues or haptics (for palpation) calibrated against physical references to further reduce cognitive load.

7.2 Closing reflection

If there is a single lesson from this research, it is that the decision to *couple* modalities is itself an engineering act. By treating optical 3D imaging and AR as complementary instruments to be composed only when composition pays for itself, the work shifts the emphasis from eye-catching demos to deployable, human-centred systems. The recommended roadmap—spanning clinical validation, regulatory rigor, and technical optimization—targets the remaining gaps so that the next iteration of this line of research can move from robust prototypes to tools that clinicians trust, institutions can maintain, and patients ultimately benefit from.

Appendix A

Publications

This research activity has led to several publications in international journals and conferences. These are summarised below. [1]

International Journals

1. **Realistic Texture Mapping of 3D Medical Models Using RGBD Camera for Mixed Reality Applications**, Aliani Cosimo, Morelli Alberto, Rossi Eva, Lombardi Sara, Civale Vincenzo Yuto, Sardini Vittoria, Verdino Flavio and Bocchi Leonardo. *Applied Sciences*, vol. 14, 2024. (Special Issue: 10), [DOI:10.3390/app14104133].
2. **Optimising Camera–ChArUco Geometry for Motion Compensation in Standing Equine CBCT**, Cosimo Aliani, Cosimo Lorenzetto Bologna, Piergiorgio Francia, and Leonardo Bocchi. *Sensors*, vol. xx, xx. [DOI:10.20944/preprints202601.1413.v1] — At the moment, this paper is still in preprints.

International Conferences and Workshops

1. **Optimizing texture representation in 3D medical models using an RGBD camera**, Aliani Cosimo, Bocchi Leonardo. "MetroInd4.0 and IoT 2024", in *Proc. of 2024 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd4.0 and IoT 2024*, Florence (Italy), 2024.
2. **Neural network-based segmentation and rendering of anatomical structures in augmented reality from tomographic images**, Alberto

[1] The author's bibliometric indices are the following: H-index = 4, total number of citations = 41 (source: Scopus on Month 2, 2026).

Morelli, Cosimo Aliani, Leonardo Bocchi. "IX Congress of the National Group of Bioengineering (GNB)", in *Proc. of Nine National Congress of Bioengineering*, Palermo (Italy), 2025.

3. **Neural network-based pose estimation and real-time tracking of ultrasound probes**, Cosimo Aliani, Alberto Morelli, Leonardo Bocchi. "47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)", Copenhagen (Denmark), 2025.

Bibliography

[1] I. Barbero-García, J. L. Lerma, P. Miranda, and Á. Marqués-Mateu, "Smartphone-based photogrammetric 3d modelling assessment by comparison with radiological medical imaging for cranial deformation analysis," *Measurement*, vol. 131, pp. 372–379, 2019.

[2] A. Haleem and M. Javaid, "3d scanning applications in medical field: a literature-based review," *Clinical Epidemiology and Global Health*, vol. 7, no. 2, pp. 199–210, 2019.

[3] M. Wells, L. N. Goldstein, T. Wells, N. Ghazi, A. Pandya, B. Furht, G. Engstrom, M. T. Jan, and R. Shih, "Total body weight estimation by 3d camera systems: Potential high-tech solutions for emergency medicine applications? a scoping review," *JACEP Open*, vol. 5, no. 5, p. e13320, 2024.

[4] A. Maiese, A. C. Manetti, C. Ciallella, and V. Fineschi, "The introduction of a new diagnostic tool in forensic pathology: Lidar sensor for 3d autopsy documentation," *Biosensors*, vol. 12, no. 2, p. 132, 2022.

[5] C.-Y. Chiu, M. Thelwell, T. Senior, S. Choppin, J. Hart, and J. Wheat, "Comparison of depth cameras for three-dimensional reconstruction in medicine," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 233, no. 9, pp. 938–947, 2019.

[6] F. Pristerà, A. Gallo, S. Fregola, A. Merola *et al.*, "Development of a biomechatronic device for motion analysis through a rgb-d camera," *Global Clinical Engineering Journal*, vol. 2, no. 3, pp. 35–44, 2020.

[7] K. A. Tychola, I. Tsimperidis, and G. A. Papakostas, "On 3d reconstruction using rgb-d cameras," *Digital*, vol. 2, no. 3, pp. 401–421, 2022.

[8] D. Filko, R. Cupec, and E. K. Nyarko, "Wound measurement by rgb-d camera," *Machine vision and applications*, vol. 29, no. 4, pp. 633–654, 2018.

[9] L. Ulrich, E. Vezzetti, S. Moos, and F. Marcolin, "Analysis of rgb-d camera technologies for supporting different facial usage scenarios," *Multimedia Tools and Applications*, vol. 79, no. 39, pp. 29 375–29 398, 2020.

[10] G. Sansoni, M. Trebeschi, and F. Docchio, "State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation," *Sensors*, vol. 9, no. 1, pp. 568–601, 2009.

[11] U. Wijenayake and S.-Y. Park, "Real-time external respiratory motion measuring technique using an rgb-d camera and principal component analysis," *Sensors*, vol. 17, no. 8, p. 1840, 2017.

[12] A. Fuster-Guilló, J. Azorin-Lopez, M. Saval-Calvo, J. M. Castillo-Zaragoza, N. Garcia-D'Urso, and R. B. Fisher, "Rgb-d-based framework to acquire, visualize and measure the human body for dietetic treatments," *Sensors*, vol. 20, no. 13, p. 3690, 2020.

[13] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 757–764.

[14] M. L. Heilig, "Stereoscopic-television apparatus for individual use," Oct. 4 1960, uS Patent 2,955,156.

[15] J. Ramalhinho, S. Yoo, T. Dowrick, B. Koo, M. Somasundaram, K. Gurusamy, D. J. Hawkes, B. Davidson, A. Blandford, and M. J. Clarkson, "The value of augmented reality in surgery—a usability study on laparoscopic liver surgery," *Medical Image Analysis*, vol. 90, p. 102943, 2023.

[16] B. Acidi, M. Ghallab, S. Cotin, E. Vibert, and N. Golse, "Augmented reality in liver surgery," *Journal of Visceral Surgery*, vol. 160, no. 2, pp. 118–126, 2023.

[17] S. Bin, S. Masood, and Y. Jung, "Virtual and augmented reality in medicine," in *Biomedical information technology*. Elsevier, 2020, pp. 673–686.

[18] E. Barcali, E. Iadanza, L. Manetti, P. Francia, C. Nardi, and L. Bocchi, "Augmented reality in surgery: a scoping review," *Applied Sciences*, vol. 12, no. 14, p. 6890, 2022.

[19] P. Dhar, T. Rocks, R. M. Samarasinghe, G. Stephenson, and C. Smith, "Augmented reality in medical education: students' experiences and learning outcomes," *Medical education online*, vol. 26, no. 1, p. 1953953, 2021.

[20] B. Felix, S. B. Kalatar, B. Moatz, C. Hofstetter, M. Karsy, R. Parr, and W. Gibby, "Augmented reality spine surgery navigation: increasing pedicle screw insertion accuracy for both open and minimally invasive spine surgeries," *Spine*, vol. 47, no. 12, pp. 865–872, 2022.

[21] F. Giannone, E. Felli, Z. Cherkaoui, P. Mascagni, and P. Pessaux, "Augmented reality and image-guided robotic liver surgery," *Cancers*, vol. 13, no. 24, p. 6268, 2021.

[22] T. Tene, D. F. Vique López, P. E. Valverde Aguirre, L. M. Orna Puente, and C. Vacacela Gomez, "Virtual reality and augmented reality in medical education: an umbrella review," *Frontiers in digital health*, vol. 6, p. 1365345, 2024.

[23] B. Puladi, M. Ooms, M. Bellgardt, M. Cesov, M. Lipprandt, S. Raith, F. Peters, S. C. Möhlhenrich, A. Prescher, F. Hölzle *et al.*, "Augmented reality-based surgery on the human cadaver using a new generation of optical head-mounted displays: development and feasibility study," *JMIR Serious Games*, vol. 10, no. 2, p. e34781, 2022.

[24] J. Hofman, P. De Backer, I. Manghi, J. Simoens, R. De Groote, H. Van Den Bossche, M. D'Hondt, T. Oosterlinck, J. Lippens, C. Van Praet *et al.*, "First-in-human real-time ai-assisted instrument deocclusion during augmented reality robotic surgery," *Healthcare Technology Letters*, vol. 11, no. 2-3, pp. 33–39, 2024.

[25] T. Morimoto, T. Kobayashi, H. Hirata, K. Otani, M. Sugimoto, M. Tsukamoto, T. Yoshihara, M. Ueno, and M. Mawatari, "Xr (extended reality: virtual reality, augmented reality, mixed reality) technology in spine medicine: status quo and quo vadis," *Journal of Clinical Medicine*, vol. 11, no. 2, p. 470, 2022.

[26] A. W. K. Yeung, A. Tosevska, E. Klager, F. Eibensteiner, D. Laxar, J. Stoyanov, M. Glisic, S. Zeiner, S. T. Kulnik, R. Crutzen *et al.*, "Virtual and augmented reality applications in medicine: analysis of the scientific literature," *Journal of medical internet research*, vol. 23, no. 2, p. e25499, 2021.

[27] P. Sun, Y. Zhao, J. Men, Z.-R. Ma, H.-Z. Jiang, C.-Y. Liu, and W. Feng, "Application of virtual and augmented reality technology in hip surgery: systematic review," *Journal of medical Internet research*, vol. 25, p. e37599, 2023.

[28] M. Eckert, J. S. Volmerg, and C. M. Friedrich, "Augmented reality in medicine: systematic and bibliographic review," *JMIR mHealth and uHealth*, vol. 7, no. 4, p. e10967, 2019.

[29] S. Nanchahal, A. Arjomandi Rad, V. Naruka, J. Chacko, G. Liu, J. Afoke, G. Miller, J. Malawana, and P. Punjabi, "Mitral valve surgery assisted by virtual and augmented reality: Cardiac surgery at the front of innovation," *Perfusion*, vol. 39, no. 2, pp. 244–255, 2024.

[30] E. Z. Barsom, M. Graafland, and M. P. Schijven, "Systematic review on the effectiveness of augmented reality applications in medical training," *Surgical endoscopy*, vol. 30, no. 10, pp. 4174–4183, 2016.

[31] C. Kamphuis, E. Barsom, M. Schijven, and N. Christoph, "Augmented reality in medical education?" *Perspectives on medical education*, vol. 3, no. 4, pp. 300–311, 2014.

[32] J.-C. Chien, J.-D. Lee, C.-W. Chang, and C.-T. Wu, "A projection-based augmented reality system for medical applications," *Applied Sciences*, vol. 12, no. 23, p. 12027, 2022.

[33] C. Kaewrat, C. Khundam, and M. Thu, "Enhancing exercise monitoring and guidance through mobile augmented reality: A comparative study of rgb and lidar," *IEEE Access*, 2024.

[34] X. Wang, S. Habert, M. Ma, C.-H. Huang, P. Fallavollita, and N. Navab, "[poster] rgb-d/c-arm calibration and application in medical augmented reality," in *2015 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2015, pp. 100–103.

[35] S. C. Lee, B. Fuerst, J. Fotouhi, M. Fischer, G. Osgood, and N. Navab, "Calibration of rgbd camera and cone-beam ct for 3d intra-operative mixed reality visualization," *International journal of computer assisted radiology and surgery*, vol. 11, no. 6, pp. 967–975, 2016.

[36] M.-L. Wu, J.-C. Chien, C.-T. Wu, and J.-D. Lee, "An augmented reality system using improved-iterative closest point algorithm for on-patient medical image visualization," *Sensors*, vol. 18, no. 8, p. 2505, 2018.

[37] [Online]. Available: https://commons.wikimedia.org/wiki/File:Intel_Realsense_depth_camera_D435.jpg

[38] [Online]. Available: https://www.01net.it/intel-realsense-l515-telecamera-lidar/

[39] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[41] Itseez, "Open source computer vision library," https://github.com/itseez/opencv, 2015.

[42] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[43] [Online]. Available: https://www.deepen.ai/blog/what-is-a-charuco-board-and-why-you-should-use-it

[44] C. Aliani, C. L. Bologna, P. Francia, and L. Bocchi, "Optimising camera–charuco geometry for motion compensation in standing equine cbct," *Preprints*, January 2026. [Online]. Available: https://doi.org/10.20944/preprints202601.1413.v1

[45] N. Otsu *et al.*, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[46] C. Aliani and L. Bocchi, "Optimizing texture representation in 3d medical models using an rgbd camera," in *2024 IEEE International Workshop on Metrology for Industry 4.0 IoT (MetroInd4.0 IoT)*, 2024, pp. 111–116.

[47] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[48] C. Aliani, A. Morelli, E. Rossi, S. Lombardi, V. Y. Civale, V. Sardini, F. Verdino, and L. Bocchi, "Realistic texture mapping of 3d medical models using rgbd camera for mixed reality applications," *Applied Sciences*, vol. 14, no. 10, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/10/4133

[49] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "Foundationpose: Unified 6d pose estimation and tracking of novel objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 868–17 879.

[50] C. Aliani, A. Morelli, and L. Bocchi, "Neural network-based pose estimation and real-time tracking of ultrasound probes," in *IEEE Engineering in Medicine and Biology Society IEEE (EMBC)*, 2025.

[51] [Online]. Available: https://github.com/mlavik1/UnityVolumeRendering

[52] [Online]. Available: https://otticauniversitaria.it/elgato-camlink-4k/