POLITECHNIKA ŚLĄSKA, WYDZIAŁ AUTOMATYKI,
ELEKTRONIKI I INFORMATYKI

DOCTORAL THESIS

Tracking Head Movements in a Flight Simulator Environment using SLAM Mechanism of Augmented Reality Goggles.

Author:
Onyeka Josephine NWOBODO

Supervisor:
Dr hab. inż. Krzysztof Adam CYRAN, Prof. PŚ

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Katedra Grafiki, Wizji Komputerowej i Systemów Cyfrowych

September 30, 2025

Abstract

Onyeka Josephine NWOBODO

Tracking Head Movements in a Flight Simulator Environment using SLAM Mechanism of Augmented Reality Goggles.

The evolving complexity of modern aviation demands continuous advancements in pilot training technologies. This dissertation addresses a critical gap in flight simulation by developing and validating an innovative augmented reality (AR)-based head-tracking system that utilizes Simultaneous Localisation and Mapping (SLAM) technology, integrated with an adaptive Kalman-Particle filter fusion. By enhancing the fidelity, responsiveness, and realism of AR flight simulators, the research makes a transformative contribution to the aviation industry, leading to improved pilot training, increased operational safety, and heightened preparedness for real-world scenarios.

The motivation for this research stems from the pressing need to overcome the limitations of existing flight simulators, which often suffer from tracking inaccuracies, latency issues, and insufficient adaptability to dynamic cockpit environments. Flight simulators are indispensable in pilot training, allowing for the practice of complex manoeuvres, emergency scenarios, and diverse flight conditions in a safe, controlled environment. However, the effectiveness of these simulators is heavily reliant on the accuracy and responsiveness of the head-tracking systems that drive the AR interfaces. Current head-tracking technologies, including marker-based systems and inertial sensors, frequently encounter challenges such as drift, occlusion, and environmental noise. These limitations can lead to misalignment between virtual and physical elements, ultimately disrupting the immersive experience and reducing the effectiveness of training. Such constraints undermine the immersive quality of AR simulations, hindering pilots from fully engaging in prolonged, realistic training exercises. This thesis addresses these challenges directly by proposing a novel SLAM-based approach tailored to the specific requirements of flight simulation environments. SLAM technology enables real-time cockpit mapping, dynamically adjusting to head movements and environmental shifts, ensuring that virtual overlays remain consistently aligned with physical cockpit elements. This dynamic recalibration enhances pilot situational awareness, allowing fluid transitions between simulated and real-world environments.

A comprehensive literature review highlights the deficiencies of existing marker and IMU-based tracking methods, establishing a foundation for integrating SLAM technologies. This research rigorously evaluates various SLAM algorithms, including ORB-SLAM, RGB-D SLAM, and Visual-Inertial SLAM (VINS). The selection and refinement of these algorithms play a crucial role in enhancing the robustness and accuracy of AR-based flight simulators. The integration of SLAM reduces latency and significantly improves tracking precision, addressing the core limitations of existing

systems. In parallel, this dissertation provides a detailed comparison between marker-based tracking and SLAM-based tracking for flight simulators. While cost-effective and straightforward, Marker-based tracking often struggles with occlusion and drift, particularly in complex, multi-angle environments such as aircraft cockpits. SLAM-based tracking, by contrast, offers superior adaptability by continuously updating the environment's spatial map. This study evaluates the performance of both approaches across diverse flight scenarios, revealing that SLAM-based tracking consistently outperforms marker-based methods in adaptability, responsiveness, and overall immersion. The comparative analysis underscores SLAM's potential to revolutionize pilot training, enabling more realistic and uninterrupted AR experiences.

Complementing spatial fidelity, the research also develops a machine-learning identification layer that continuously and implicitly recognizes the trainee and their current task from routine AR interaction. A multimodal, multi-branch LSTM fuses time-synchronized eye-gaze, head-motion, and hand-gesture streams to derive stable behavioural biometrics for continuous authentication and task recognition on-device. This layer replaces cumbersome keyboard/PIN login on shared headsets, strengthens privacy and access control, and enables task-contingent personalization so overlays, checklists, and assessments are bound to the correct user and adapted in real time without interrupting training.

Extending adaptivity beyond identification, the dissertation contributes a light-weight deep-learning framework that predicts memory-retention state using non-invasive behavioural signals captured by HoloLens 2: gaze duration, interaction counts, revisit counts, and head-movement stability, augmented with content category embeddings. By inferring states such as strong recall or cognitive overload, the framework supports real-time pacing, targeted repetition, and scaffolded guidance, thereby closing the loop between how trainees interact and what they retain and advancing the goal of low-load, effective training.

The methodological framework of this research encompasses extensive experimental setups and comparative analyses of diverse filtering approaches, including Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Ensemble Kalman Filter (EnKF), and Sequential Importance Resampling (SIR) particle filters. The innovative fusion of Kalman and particle filters is shown to leverage the complementary strengths of both approaches, addressing the nonlinear dynamics inherent to AR head tracking in flight simulation. This hybrid approach enhances the overall resilience and accuracy of the tracking system, resulting in superior stability during flight simulation exercises. Experimental validations demonstrate that this adaptive fusion method consistently reduces the root mean square error (RMSE) across spatial tracking metrics, resulting in a notable increase in stability and precision.

Moreover, the research examines the impact of these technological advancements on the efficacy of pilot training. One of the most significant findings is the reduction in latency and motion sickness experienced during extended simulation sessions. Traditional flight simulators, which rely heavily on marker-based tracking, often suffer from delayed responses that can induce motion sickness in pilots. The refined tracking precision enabled by SLAM minimizes cognitive dissonance, allowing pilots to remain engaged in immersive training environments for more extended periods without

discomfort. This breakthrough enhances training outcomes and promotes the cost-effective scaling of AR-based flight simulators for widespread adoption in commercial and military aviation sectors.

The thesis also contributes a novel adaptation of Fitts' Law for AR environments, offering a theoretical framework for evaluating user interactions in three-dimensional space. By incorporating head movements and spatial orientation into predictive models, this adaptation reduces interaction task difficulty by approximately 40%, thereby optimizing AR interface designs for an enhanced user experience. This insight paves the way for more intuitive cockpit interfaces and ergonomic AR system layouts, further reinforcing the simulator's value in comprehensive pilot training programs.

Beyond aviation, the platform's transferability is illustrated through an AR Mach–Zehnder Interferometer learning module that applies the same spatial, identity-aware, and cognition-adaptive design principles to concept-dense STEM content. The module converts abstract quantum-optics constructs into SLAM-anchored, manipulable representations of optical elements and beam paths and couples guided exploration with reconstruction to scaffold schema formation, establishing a reusable template for translating abstract technical knowledge into situated, task-guided interaction.

This dissertation represents a pivotal advancement in AR flight simulation by addressing core issues in head tracking, spatial mapping, and real-time interaction. The adaptive Kalman-Particle filter fusion method, combined with SLAM-based tracking, establishes new performance benchmarks for AR systems in aviation. This research lays the groundwork for future AR developments in flight simulation. It extends its applicability to other domains requiring high-fidelity spatial tracking, such as unmanned aerial vehicle (UAV) piloting and air traffic management. By bridging the technological divide between simulated and real-world flight environments, this dissertation underscores the transformative potential of AR in aviation, ushering in a new era of immersive, responsive, and effective pilot training solutions.

Acknowledgements

First and foremost, I express my profound gratitude to God for the strength, wisdom, and knowledge to undertake and complete this journey. In moments of doubt and fatigue, that grace sustained me and illuminated the path forward.

To my parents, whose love, sacrifices, and unwavering belief in me made this possible; to my sibling, for constant encouragement and patience; and to my friends, for their understanding, kindness, and steadfast support, thank you. Your presence has been a source of courage and joy.

I am deeply grateful to my supervisor for unwavering guidance, generous mentorship, and continual support from the outset of this project. Your high standards, insight, and trust challenged me to do my best work and to develop as an independent scholar. I also thank the members of my dissertation committee for their rigorous feedback and insightful suggestions. Your perspectives strengthened the arguments and improved the scholarship presented in this dissertation.

This project benefited enormously from the collegiality of my department and from financial support that allowed me to focus on my studies. To my colleagues, thank you for your camaraderie, stimulating discussions, and the countless ways you helped, seen and unseen. Working alongside you has been a privilege.

I am grateful to the collaborators and participants who took part in the experiments and contributed data; this dissertation would not have been possible without your cooperation and trust.

Finally, to everyone who walked with me along the way; my supervisor, family, friends, mentors, and colleagues, thank you. Your encouragement and example sustained me throughout this journey, and I remain deeply grateful.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	**A**rtificial **I**ntelligence
AKPF	**A**daptive **K**alman-**P**article **F**ilter
ANOVA	**A**nalysis of **V**ariance
APF	**A**uxiliary **P**article **F**ilter
AR	**A**ugmented **R**eality
BA	**B**undle **A**djustment
BLEU	**B**ilingual **E**valuation **U**nderstudy (score)
CNN	**C**onvolutional **N**eural **N**etwork
CPU	**C**entral **P**rocessing **U**nit
CSV	**C**omma-**S**eparated **V**alues
DNN	**D**eep **N**eural **N**etwork
DoF / 6DoF	(Six) **D**egrees of **F**reedom
EEG	**E**lectroencephalography
EKF	**E**xtended **K**alman **F**ilter
EnKF	**E**nsemble **K**alman **F**ilter
FFS	**F**ull-**F**light **S**imulator
FOV	**F**ield of **V**iew
fps	**F**rames **P**er **S**econd
GNSS	**G**lobal **N**avigation **S**atellite **S**ystem
GPU	**G**raphics **P**rocessing **U**nit
GUI	**G**raphical **U**ser **I**nterface
HCI	**H**uman-**C**omputer **I**nteraction
HMD	**H**ead-**M**ounted **D**isplay
HPU	**H**olographic **P**rocessing **U**nit
HUD	**H**eads-**U**p **D**isplay
IAICP	**I**ntensity-**A**ssisted **I**terative **C**losest **P**oint
ICP	**I**terative **C**losest **P**oint
IMU	**I**nertial **M**easurement **U**nit
IR	**I**nfra**r**ed
IS	**I**nteractive **S**imulation
LSD-SLAM	**L**arge-**S**cale **D**irect **S**LAM
LSTM	**L**ong **S**hort-**T**erm **M**emory (network)
MAE	**M**ean **A**bsolute **E**rror
MR	**M**ixed **R**eality
MSE	**M**ean **S**quared **E**rror
MTL	**M**edial **T**emporal **L**obe
MZI	**M**ach–**Z**ehnder **I**nterferometer
ORB-SLAM	**O**riented **F**AST and **R**otated **BRIEF** based **SLAM** algorithm

PAC	**P**hase-**A**mplitude **C**oupling
PnP	**P**erspective-**n**-**P**oint
ReLU	**R**ectified **L**inear **U**nit
RGB	**R**ed, **G**reen, **B**lue
RGB-D	**RGB** plus **D**epth
RGB-D SLAM	**RGB-D**epth based **SLAM** algorithm
RM	**R**esearch **M**ode (HoloLens 2)
RMSE	**R**oot **M**ean **S**quare **E**rror
RNN	**R**ecurrent **N**eural **N**etwork
SDK	**S**oftware **D**evelopment **K**it
SIR (PF)	**S**equential **I**mportance **R**esampling (**P**article **F**ilter)
SLAM	**S**imultaneous **L**ocalization and **M**apping
SVO	**S**emi-**D**irect **V**isual **O**dometry
t-SNE	**t**-**D**istributed **S**tochastic **N**eighbor **E**mbedding
ToF	**T**ime-of-**F**light (depth sensing)
UPF	**U**nscented **P**article **F**ilter
USB-C	**U**niversal **S**erial **B**us Type-**C**
VINS	**V**isual-**I**nertial **SLAM**
VR	**V**irtual **R**eality
Wi-Fi	**W**ireless **F**idelity
YOLO	**Y**ou **O**nly **L**ook **O**nce (object detection)

1 Introduction

1.1 Background

The role of flight simulators in pilot training has evolved profoundly over the last century, establishing them as an essential component of modern aviation [1], [2]. The origins of flight simulation technology can be traced to the early 20th century when rudimentary devices were first designed to replicate basic flight dynamics, providing pilots with foundational training in a controlled environment [3]. The development of these early simulation devices was motivated by the necessity to familiarize pilots with flight principles and procedures before they operated actual aircraft, a task that was both costly and dangerous to conduct exclusively in real flight conditions [4]. A pivotal advancement in flight simulation was achieved with the introduction of the Link Trainer in the 1920s, illustrated in Figure 1.1, developed by Edwin Link. This mechanical simulator, often called the "Blue Box," represented a groundbreaking solution for training pilots, especially in the challenging area of instrument flight. Unlike previous methods, which primarily relied on direct flight experience, the Link Trainer provided a replicable, safe environment where pilots could practice flight techniques critical for low-visibility or night-time navigation. The Link Trainer's utility was especially evident during World War II when many pilots needed to be trained rapidly and effectively. It became widely used by military aviation programs, providing structured training on essential skills such as instrument navigation, handling simulated malfunctions, and responding to adverse weather conditions. The Link Trainer's impact on pilot training was significant, as it not only improved the safety and efficiency of training but also set a precedent for future developments in aviation training technology, laying the foundation for the entire field of flight simulation.

The transition from mechanical to electronic flight simulators marked the next primary phase in this evolution, reflecting the increasing complexity of aviation systems and the demand for higher fidelity in training. In the latter half of the 20th century, advances in electronics and computing enabled the creation of digital simulation systems that could replicate real-world flight conditions with unprecedented precision [6]. This shift allowed for more intricate training scenarios and more extraordinary adaptability in simulator design, allowing training modules to incorporate various aircraft models, weather conditions, and emergencies [7]. As electronic simulators advanced, they included data-driven models, improving flight dynamics' realism and visual displays that provided immersive cockpit experiences. These advancements allowed pilots to experience a wide range of realistic training scenarios, helping to develop critical skills without the risks associated with live flight [8]. A visual representation of a modern flight simulator setup is shown in Figure 1.2. Such sets-ups have evolved significantly, from simple mechanical controls to highly advanced systems with motion-cueing

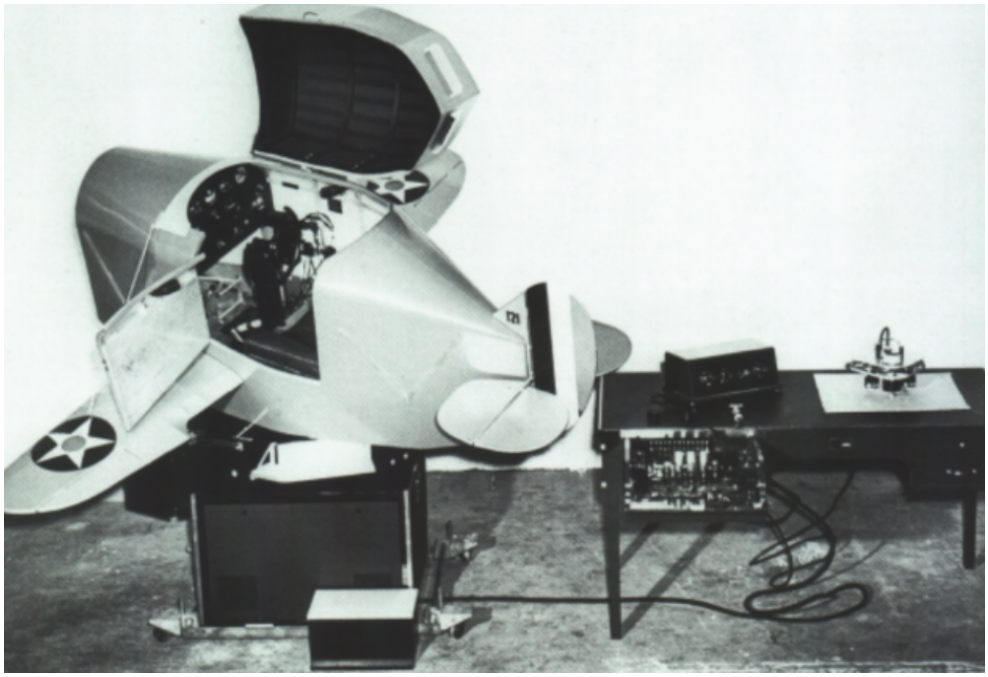


FIGURE 1.1: Link Trainer, one of the first mechanical flight simulators developed by Edwin Link in the 1920s. This device played a crucial role in pilot training during World War II. (Source: [5]).

capabilities and complex visual displays, all aimed at creating as realistic environment as possible. This transformation underscores the indispensable role of simulators in pilot training and highlights their ongoing importance in preparing pilots to navigate modern aviation's increasingly complex and congested airspace.

In recent decades, flight simulation technology has continued to advance, incorporating highly sophisticated systems such as full-motion platforms, high-resolution visual displays, and integration with real-time flight data. These components contribute to creating an environment that closely mimics actual flight, thereby offering a highly effective means of training pilots to handle complex airspace environments and the demands of modern aviation. For example, motion platforms used in modern simulators replicate the physical sensations of flight, such as turbulence and takeoff acceleration, improving the realism and immersion of the training experience [9]. Visual systems have also been significantly improved, allowing for high-definition 360-degree views of the simulated environment, including variable weather, day-night cycles, and dynamic landscapes. The accuracy and immersive quality of these systems are essential for developing pilot spatial awareness and decision-making skills in situations that closely resemble real-world conditions.

Today, flight simulators are integral to pilot training programs in various applications, including pilot certification, emergency response drills, commercial aviation, and military operations, [10], [11], [12]. They offer invaluable advantages by allowing pilots to practice and refine their skills in an environment that removes the inherent risks of live flight while preserving the fidelity needed for effective learning. This safe and controlled environment reduces the potential hazards of training in an actual aircraft.

FIGURE 1.2: Modern Flight Simulator Setup. Such simulators provide immersive environments for advanced pilot training. (Source: Wright-BroS Project).

It allows for the repeated practice of rare but critical emergency procedures, enhancing the safety and effectiveness of training programs. In addition, simulators support a range of specialized training needs, from handling adverse weather to managing in-flight system failures, which are essential for the comprehensive development of piloting skills.

1.1.1 Emerging Technologies in Flight Simulation

As flight simulators evolve, new technologies, such as augmented reality (AR) and virtual reality (VR), are increasingly integrated into simulation environments, providing new dimensions of realism and interactivity. One promising area of development is the use of AR goggles equipped with Simultaneous Localization and Mapping (SLAM) technology to track pilot head movements in flight simulators. This technology bridges the gap between virtual and physical environments, enhancing immersion and creating more natural and intuitive training scenarios. SLAM, a core technology in robotics and computer vision, allows AR goggles to continuously map the surrounding environment while simultaneously tracking the user's position within that space [13]. In flight simulators, SLAM mechanisms can track pilot head movements, enabling the system to render the correct perspective and respond dynamically to gaze and head orientation shifts. This capability is critical for fostering situational awareness and improving pilot reaction times in high-stress scenarios. By integrating head movement tracking, the simulator environment can adjust cockpit displays and

external views to reflect the pilot's natural focus, enhancing depth perception and spatial awareness. The AR environment augmented by SLAM can overlay critical flight data, navigational cues, and system warnings directly within the pilot's field of view, reducing the need for pilots to divert their attention to multiple instrumentation panels [14]. This overlay capability can significantly improve multitasking efficiency and situational awareness, addressing key challenges in modern flight operations where rapid information processing is essential. Moreover, SLAM technology enables real-time calibration of flight environments, adapting dynamically to external conditions simulated within the system [15]. For example, turbulence effects, weather conditions, and emergency scenarios can be reflected more precisely as head-tracking data feeds into the simulation, allowing the AR interface to guide pilots more effectively during crisis management scenarios. The application of SLAM technology in AR goggles has the potential to transform traditional flight simulation by creating a highly immersive, responsive environment that more accurately mimics real-world flight conditions [16]. This approach also addresses the limitations of conventional static displays, which often lack the depth and interactivity required to accurately simulate complex, real-time decision-making processes. As a result, AR-enabled simulators with SLAM tracking provide a more holistic training experience, allowing pilots to engage with dynamic elements of flight training, such as formation flying [17], aerial refuelling, [18], and emergency manoeuvring [19]. Furthermore, the development of AR and SLAM-integrated flight simulators aligns closely with the broader goal of enhancing pilot proficiency while reducing costs associated with live flight training [20]. By simulating high-risk scenarios in a controlled, immersive environment, AR-based simulators mitigate the risks involved in live exercises and allow for extensive practice of rare emergency procedures critical to pilot preparedness. The continued evolution of flight simulators driven by AR and SLAM technologies represents a significant step forward in aviation training. As the aviation industry seeks to address increasing demands for safety, efficiency, and operational readiness, these innovations promise to play a vital role in shaping the future of pilot education and flight proficiency. In addition to spatial fidelity, effective training also requires systems that adapt to the individual using them. Modeling user-specific behavior—such as patterns of gaze, head motion, and interaction—enables the simulator to deliver more personalized, responsive, and realistic training experiences. As part of this broader goal, this thesis investigates AI-driven techniques for identifying and modeling individual users based on their behavioral data during AR interactions.

1.1.2 Problem Statement

Flight simulators are indispensable in pilot training, enabling realistic, safe environments where pilots can practice under controlled scenarios, including diverse and complex conditions like air density variations, turbulence, and wind shear. As aviation systems and airspace operations grow increasingly complex, augmented reality (AR) has emerged as a promising innovation to enhance flight simulator fidelity by overlaying digital information onto physical cockpit elements, creating an immersive training experience [21]. Despite AR's potential, achieving accurate virtual-physical alignment and maintaining situational awareness remain significant challenges, primarily due to the limitations of current head-tracking systems [22]. As aviation technology advances,

pilot training must evolve to mirror the increasing intricacies of modern cockpit environments and flight conditions. This need has propelled the integration of AR into flight simulation, offering opportunities to enrich pilot experiences by overlaying vital flight data and navigational cues directly onto their field of view. However, existing head-tracking systems have yet to achieve the precision required to blend virtual and physical elements seamlessly; a critical shortfall undermines AR's transformative potential [23]. The misalignment of AR overlays detracts from the immersive experience and impedes the development of spatial awareness, a core competency in pilot training [24]. AR elements may lag or deviate without accurate and responsive head-tracking, potentially leading to cognitive dissonance and impairing the pilot's ability to respond effectively to simulated flight conditions.

Head-tracking technologies in augmented reality (AR)-based simulators, including infrared (IR)-based and inertial measurement unit (IMU)-based systems, face critical limitations. IR-based systems are prone to interference from variable lighting within the cockpit, leading to inaccuracies caused by reflections or ambient light sources. For instance, research has shown that IR-based head-tracking systems can experience performance degradation due to ambient lighting conditions and reflective surfaces, which can introduce noise and reduce tracking accuracy [23]. While providing continuous tracking, IMU-based systems are affected by sensor drift, which causes cumulative inaccuracies over time [25]. These issues necessitate frequent recalibration, adding computational overhead and diminishing responsiveness during high-stakes training scenarios. Consequently, these limitations often lead to misalignment between virtual overlays and physical cockpit components, disrupting the immersive experience and reducing the training's overall effectiveness. When pilots are forced to correct system errors rather than focus on situational awareness and decision-making, the simulator's value as a training tool is compromised [26].

Simultaneous Localization and Mapping (SLAM) has shown potential to solve these challenges. SLAM technology, commonly used in robotics and augmented reality, enables simultaneous tracking and environmental mapping in real-time, adapting dynamically to the user's movement and the surrounding environment [27]. Within a flight simulation context, SLAM could compensate for the limitations of IR and IMU systems by providing more accurate tracking, adjusting to variable lighting, countering sensor drift, and offering real-time error correction, thereby supporting precise alignment between virtual elements and physical cockpit components [13]. The robustness of SLAM lies in its ability to create a continuously updated map of the cockpit environment, allowing AR systems to recalibrate rapidly in response to movement, thus maintaining seamless virtual-physical interaction [28]. This technology also enhances the accuracy of perspective shifts, ensuring that digital overlays reflect realistic changes in pilot head position, a crucial factor in replicating real-life cockpit dynamics [29]. However, implementing SLAM in AR systems for flight simulation introduces its technical challenges, including managing nonlinear sensor data, minimizing data association errors, and ensuring robust performance within an aircraft cockpit confined and intricate environment [30]. The cockpit's reflective surfaces and narrow space introduce complexities that necessitate sophisticated SLAM algorithms capable of handling occlusions and dynamic environments [31]. Addressing these technical hurdles is essential for achieving the high-fidelity simulation environment

required for practical pilot training.

This dissertation investigates integrating SLAM technology into AR goggles to improve head movement tracking in flight simulators. The goal is to create a more accurate, responsive, and immersive AR system that reduces cognitive load and enhances pilot training effectiveness. By addressing current technical limitations, this research seeks to demonstrate how SLAM can significantly enhance the quality and realism of flight simulation, ultimately contributing to better pilot preparedness, training efficiency, and aviation safety. This study aims to establish a foundation for future AR innovations in aviation, expanding the potential for real-time pilot-assist systems and redefining the scope of simulator-based training in line with the evolving demands of contemporary airspace operations.

1.1.3 Motivation

The increasing complexity of modern aviation systems, coupled with progressively stringent safety regulations [32], underscores the necessity for highly realistic and effective pilot training solutions. Immersive simulation technologies address this demand by narrowing the gap between theoretical instruction and practical experience. Advanced flight simulators, for example, can replicate high-risk scenarios—such as engine failures or severe weather conditions—within a controlled and risk-free environment, enabling pilots to develop critical decision-making skills and operational readiness without endangering personnel or equipment. Such synthetic training is also far more cost-effective than flying an actual aircraft [33], [34], enabling trainees to accumulate hours and practice rare scenarios without incurring prohibitive fuel or maintenance expenses. Indeed, researchers emphasize that "immersive, technology-driven training environments have the potential to address modern aviation demands better and enhance pilot proficiency" [35]. By using simulators to replicate complex traffic patterns, weather conditions, and emergency procedures with high fidelity, pilot programs can ensure that trainees develop the requisite skills to handle congested airspace and unexpected contingencies before they ever take to the air.

Augmented and virtual reality systems are now extending this realism even further by superimposing interactive 3D overlays onto cockpit environments. For instance, AR/VR headsets can present pilots with digital cockpit instruments, navigational cues, and environmental hazards in their field of view, so that users literally "touch" and manipulate virtual controls as if in a real aircraft [36]. In that system, spatial computing and high-resolution visuals gave pilots the sensation of sitting in a cockpit and operating all switches and dials, yet on a portable headset. Similarly, AR can transform a static flight checklist into an interactive learning tool: Pilots can view a control and instantly see its function or correct procedure, thus reinforcing muscle memory and situational understanding [37]. These immersive technologies have shown promise in making complex procedures more intuitive and self-paced. By combining lifelike visuals with continuous feedback, AR/VR trainers can accelerate skill acquisition, while dramatically reducing training costs (simulators and AR devices can be reused endlessly at a fraction of the cost of actual aircraft time) [20].

Despite these advances, technical limitations in current tracking systems continue to hinder full immersion. Conventional head-tracking methods (optical or inertial) can lose precision due to occlusions, lighting changes, or sensor drift, resulting in

virtual overlays that lag or misalign. Such misregistrations break the illusion of reality and can affect pilot focus. This is a serious concern given that loss of situational awareness is the basis of up to 85% of general aviation accidents [32]. To overcome this, researchers are integrating Simultaneous Localization and Mapping (SLAM) into AR flight training. SLAM uses cameras and inertial sensors to build a live map of the cockpit environment and track the user's head pose in real time. In practice, SLAM can continuously recalibrate the AR display so that the digital gauges and external visuals remain precisely aligned with the physical cockpit as the pilot moves. Early studies report that SLAM-based trackers can operate in real time with very high accuracy; for example, the LSD SLAM method (Large Scale Direct Simultaneous Localization and Mapping) is noted for its 'real-time performance' and ability to 'precisely follow the movement of the simulator'. In concrete terms, AR equipped with SLAM (such as Microsoft HoloLens) has been shown to produce a more accurate and stable AR experience by persistently mapping and registering the environment [16]. As a result, pilots can maintain continuous immersion even during rapid head turns or turbulent maneuvers, preserving situational awareness, and ensuring that virtual training cues remain reliable and believable.

By merging AR with robust SLAM tracking, the next generation of simulators aims to deliver unparalleled training fidelity. Such systems promise to let pilots train 'anytime from anywhere' with complete immersion in the cockpit, reducing the need for expensive flight hours. Industry surveys indicate that integrating AR into simulator devices significantly reduces costs while improving learning outcomes [35]. Looking ahead, SLAM-enhanced AR could extend beyond pilot cockpit training to other domains, such as air traffic controller simulators, unmanned drone operator training, and even spacecraft flight drills, aligning with the aviation industry's goals of greater safety and efficiency. In summary, immersive AR/VR training platforms augmented by SLAM have the potential to make pilot education more realistic, adaptable, and cost-effective than ever before.

1.1.4 Research Questions and Objective

This thesis delves into addressing pivotal challenges in contemporary aviation training, focusing on enhancing pilot training environments through the implementation of advanced augmented reality (AR) technologies. Specifically, this research investigates the potential of integrating Simultaneous Localization and Mapping (SLAM) technology with AR goggles to significantly elevate AR-based flight simulators' fidelity, accuracy, and responsiveness. As the complexity of aviation systems continues to evolve and the aviation industry faces mounting pressure to mitigate human error, the development of immersive and precise training environments has emerged as an essential component of pilot preparation. SLAM-based AR goggles have the transformative potential to refine head-tracking precision, facilitating adaptive and dynamic simulation scenarios. Such advancements hold promise for enhancing pilot readiness by aligning training methodologies more closely with the practical demands encountered in real-world aviation operations [38].

The aviation industry is increasingly confronted with the need to manage the multifaceted challenges of congested airspace, rapidly evolving cockpit technologies, and stringent safety regulations. These factors necessitate the creation and deployment

FIGURE 1.3: Augmented Flight Simulator (Source: WrightBroS Project).

of advanced training platforms capable of replicating real-world flight conditions with high accuracy and responsiveness. Although existing flight simulators demonstrate considerable technological sophistication, they frequently fall short in terms of providing the dynamic adaptability required to effectively replicate sudden environmental changes, emergency protocols, and intricate navigational scenarios. Augmented reality systems, when augmented with SLAM technology, can bridge this critical gap by enabling interactive, real-time overlays that fluidly respond to pilot movements and evolving cockpit configurations. By fostering a more immersive and responsive simulation environment, such systems have the potential to redefine the scope and efficacy of pilot training [13]. This study investigates how these technological advancements can revolutionize pilot training by providing intuitive, accurate, and immersive simulation environments. To guide this exploration, the research is structured around the following primary research questions:

- How can SLAM-based AR goggles reduce latency in head tracking within flight simulators compared to traditional tracking methods, thereby enhancing real-time responsiveness and immersion?
- What is the optimal SLAM algorithm for achieving reliable, real-time head tracking in dynamic flight simulator environments while minimizing drift and maintaining accuracy?
- How does improved head-tracking fidelity in AR-based flight simulators impact pilot situational awareness, decision-making, and overall training effectiveness?

To comprehensively address these research questions, this study undertakes a thorough investigation into the performance enhancements facilitated by AR-SLAM

technologies, with a particular focus on latency reduction as a key determinant of real-time interaction quality. Given that flight scenarios often necessitate split-second decision-making and rapid responses under high-pressure conditions, even marginal tracking delays can undermine pilot performance and detract from the overall efficacy of training exercises. As such, this research places a pronounced emphasis on evaluating various SLAM algorithms to ascertain their effectiveness in minimizing drift and maintaining precise spatial alignment across diverse and complex simulated flight conditions [16], [39]. Through experimental testing and comparative analysis, the study endeavours to elucidate the relative strengths and limitations inherent to different SLAM implementations, ultimately providing a comprehensive assessment of their viability for large-scale integration into aviation training programs. The specific objectives underpinning this research are articulated as follows:

- To identify and evaluate practical approaches for improving SLAM mechanisms in AR applications for flight simulation, focusing on achieving high accuracy, real-time performance, and robustness in tracking.
- To address and advance current SLAM technologies, particularly in rapid initialization, accurate scaling, and resilience against jitter, rotational motion, and cockpit lighting variations.
- To enhance SLAM techniques for AR applications in aviation by developing real-time, markerless tracking methods that improve training fidelity, reduce cognitive load, and contribute to effective, immersive pilot training environments.

This research aspires to significantly contribute to advancing aviation training technologies, offering novel insights and methodologies that can enhance the fidelity and overall effectiveness of AR-based flight simulators. By integrating SLAM-based tracking improvements, the study seeks to bridge gaps in simulator realism, laying the groundwork for more comprehensive and safety-oriented pilot training practices. Furthermore, the anticipated outcomes of this research are expected to extend beyond the immediate scope of pilot training. The findings also inform developing and designing next-generation AR-assisted systems for air traffic management, unmanned aerial vehicle (UAV) operations, and cockpit interface enhancements. This underscores the broader implications and transformative potential of SLAM-driven AR technologies across multiple facets of the aviation industry. The insights derived from this study are poised to address pressing challenges in aviation training while simultaneously serving as a foundational cornerstone for future innovations, solidifying the role of AR-SLAM as a critical component in the evolution of flight simulation platforms.

1.2 Methodology

1.2.1 Overview

This study implements a comprehensive methodological framework that integrates Simultaneous Localization and Mapping (SLAM) with augmented reality (AR) head-mounted displays (HMDs) to facilitate head-tracking precision in flight simulation environments. The approach aims to mitigate common limitations in conventional

tracking systems, such as drift, latency, and interference from cockpit lighting, by deploying a hybrid estimation model that combines Kalman and particle filtering techniques for robust sensor fusion.

The methodological approach is structured across four core phases aligned with the research objectives: (1) SLAM algorithm selection and visual-inertial feature extraction, (2) system calibration and integration of the AR simulator, and (3) algorithmic optimization using an adaptive Kalman-particle filter and machine learning techniques. The fourth phase explores the practical applications of the SLAM-AR system in diverse simulation scenarios, evaluating its impact on situational awareness, cognitive load, and training realism. Optimization is employed to fine-tune model hyperparameters, ensuring predictive reliability across use cases. This dual-layered approach, combining model-based filtering with data-driven learning, aims to enhance pilot readiness and aviation safety by delivering immersive, adaptive, and cognitively aligned training systems.

1.2.2 Research Design

A mixed-methods approach was employed, integrating quantitative performance evaluations and qualitative user feedback to validate the efficacy of SLAM-based tracking systems. This dual analysis enabled comprehensive assessments of the system's technical performance while capturing subjective user experiences during interaction with AR-based cockpit simulators. The research commenced in the following phases:

Literature review (Chapter 2)

To examine existing head-tracking technologies and identify limitations inherent in marker-based systems, providing the rationale for adopting SLAM-based approaches in aviation simulation.

SLAM Implementation and Feature Extraction (Chapter 3)

To achieve real-time localization and stable AR overlays, the system integrated multiple SLAM algorithms (ORB-SLAM, RGB-D SLAM, and VINS) into a hybrid visual-inertial framework. The Microsoft HoloLens 2 provided spatial data via its embedded depth camera and IMU sensors. Feature mapping of cockpit elements enabled the precise spatial anchoring of virtual instruments, eliminating the reliance on fiducial markers and thereby enhancing training flexibility and reducing environmental constraints.

Tracking Calibration and Comparative Evaluation (Chapter 4)

To ensure precise alignment between virtual overlays and physical cockpit components, a checkerboard calibration protocol was implemented to extract intrinsic and extrinsic parameters of the HoloLens 2's depth and RGB cameras. This calibration minimized spatial registration errors, enhancing the fidelity of AR overlays during head movement. Following calibration, a controlled experimental framework was established using a lab-based cockpit replica to simulate realistic flight conditions. A

comparative evaluation was conducted between marker-based and SLAM-based tracking approaches. Quantitative performance indicators—including localization accuracy, latency, and spatial drift—were collected to assess robustness and responsiveness.

Adapting Fitts' Law for AR Interaction (Chapter 5)

To evaluate and improve the usability of AR interfaces in the cockpit simulator, the research adapted principles from Fitts' Law. Fitts' Law, which predicts the time required to move to a target area rapidly, is a classic model in human-computer interaction for evaluating pointing tasks. In this AR context, thirty participants performed target acquisition tasks with virtual cockpit controls placed at varying distances (from about 0.3 up to 6 meters away). The tasks were designed to mimic reaching or gaze-selecting instruments in a cockpit. For each participant, metrics were recorded to quantify task difficulty and performance, including: Acquisition Time, Error Rate, and Movement Efficiency.

Algorithmic Development: Adaptive Kalman-Particle Filter (AKPF) (Chapter 6)

This outlines the development of the Adaptive Kalman-Particle Filter (AKPF), a hybrid fusion model that dynamically balances Kalman filtering for linear dynamics and particle filtering for nonlinear conditions. The adaptive tuning mechanism monitored sensor noise and motion variance to ensure continuous, low-latency, high-fidelity tracking across diverse cockpit scenarios. Bayesian optimization was employed to determine optimal hyperparameters under simulated stress conditions.

User Identification via Deep Learning in AR Interaction (Chapter 7)

To enable personalized and secure interactions in AR-based flight simulation, this chapter introduces a behavioral biometric system for implicit user identification using multimodal sensor data. Built on the Microsoft HoloLens 2, the framework collects gaze trajectories, head movements, and hand gestures during normal interactions. A multi-branch Long Short-Term Memory (LSTM) neural network is used to learn temporal behavioral signatures for each user. This machine learning approach allows continuous user identification without explicit input or calibration. The integration of this system into AR simulators supports identity-aware interaction, personalization, and user continuity during immersive training tasks.

AI-Driven Behavioural Modelling and Memory Prediction (Chapter 8)

Leveraging AR interaction data—including gaze patterns, revisit frequency, and head motion—deep learning models were trained to predict user memory retention and cognitive states. The models classified learners into states such as Strong Recall or Cognitive Overload, enabling adaptive AR content delivery. This approach fostered real-time personalisation of training experiences based on user behaviour and engagement metrics.

Immersive Conceptual Learning via AR Visualization (Chapter 9)

The framework's generalizability was tested in a domain transfer case involving quantum mechanics education. An AR-based simulation of the Mach-Zehnder Interferometer (MZI) allowed learners to explore abstract quantum phenomena interactively. Results showed a 45.45% post-test improvement in conceptual scores and a 70% reconstruction success rate, validating the pedagogical efficacy of immersive, spatially anchored AR instruction beyond aviation training.

Data Collection and Evaluation

The evaluation of the SLAM-based AR tracking system involved a structured process of collecting and analyzing data across multiple experimental phases. This section outlines the methods used to measure system performance and reliability, ensuring a comprehensive assessment of the tracking system's capabilities in simulated flight environments.

Key Evaluation Metrics

The system's performance was measured using the following criteria:

- **Localization Accuracy** – To assess the alignment between virtual AR overlays and their corresponding physical cockpit components, precise measurements were taken throughout the experiments. This ensured the system's ability to consistently project AR elements in the correct spatial locations.
- **Latency** – Time delay between pilot head movement and the AR system's visual response was recorded using real-time tracking tools. Measuring latency provided insight into system responsiveness and ensured minimal delay to enhance immersion.
- **Robustness** – The system was tested under varying environmental conditions, such as cockpit lighting changes and sensor drift. Repeated tests were conducted to evaluate how the system maintained performance in dynamic or unpredictable conditions.
- **Immersion and User Feedback** – Participants provided qualitative feedback regarding their experience using the AR system, with particular focus on spatial awareness, ease of interaction, and perceived realism during simulated cockpit tasks.
- **User Cognitive Load** – Implicitly mentioned, measured via task performance and user feedback.
- **Task Performance and Training Effectiveness** – Using metrics adapted from Fitts' Law to quantify speed and accuracy in gaze-based interactions

Statistical Analysis

Quantitative data collected during each experiment were processed and analyzed using statistical methods, ensuring objective system performance evaluation. The analysis involved:

- **Mean Accuracy Rates** – Calculated to determine the overall precision of the SLAM-based tracking system.
- **Standard Deviation** – Measured to assess data variability and system consistency across different test conditions.
- **Confidence Intervals** – Applied to estimate the reliability and significance of observed improvements in head tracking and AR alignment.

A comparative analysis was performed to benchmark the SLAM-based system against traditional head-tracking methods, including IMU-based and infrared systems. This comparison demonstrated the SLAM system's ability to reduce latency and improve spatial accuracy, highlighting its advantages in flight simulation applications. This study employs a rigorous methodological framework that integrates SLAM technology and refines head-tracking systems through iterative experimentation. The methodology supports the advancement of aviation training environments, enhances pilot situational awareness, and improves flight safety.

1.3 Significance and Contributions

This research significantly contributes to AR-based flight simulation and real-time localisation by developing and applying SLAM-enhanced tracking technology in aviation training simulators. By integrating Simultaneous Localisation and Mapping (SLAM) technology with AR goggles, this work advances markerless tracking systems, eliminating the need for fixed points and allowing for more dynamic, accurate, and real-time interactions. These innovations directly benefit pilot training by improving simulator fidelity, enabling pilots to experience more immersive training scenarios that better prepare them for real-world conditions [40].

The significance of this research lies in its ability to address critical limitations in current flight simulator technologies. Traditional head-tracking systems rely heavily on fixed markers or external sensors, which often limit movement freedom and introduce latency or inaccuracies under dynamic flight conditions. SLAM-based AR solutions circumvent these challenges by continuously mapping the cockpit environment and adapting in real-time to the pilot's head movements, without the constraints of static reference points. This dynamic recalibration enhances the realism of simulated flight environments. It ensures greater precision during complex training scenarios such as low-visibility navigation, emergency response drills, and multi-aircraft coordination [13].

One of the primary contributions of this research is the reduction of latency and improvement of tracking accuracy in head-tracking systems, which are crucial for high-fidelity training environments such as flight simulators. SLAM-based AR goggles provide pilots with realistic, immersive experiences that are essential for practicing

complex maneuvers with precision. This study provides empirical evidence demonstrating that real-time SLAM-based tracking enhances situational awareness, reduces cognitive load, and improves pilot proficiency and safety. By addressing the core challenges of latency and drift, the research enables flight simulators to deliver training that closely mirrors the dynamic nature of real-world flight operations, fostering quicker response times and sharper decision-making abilities [41].

Beyond aviation, this study's contributions extend to various industries reliant on immersive simulation technologies. Healthcare is a prime beneficiary, as VR-based surgical simulations rely on precise tracking to accurately mimic delicate procedures. SLAM-enabled tracking ensures accurate interaction with virtual objects, allowing medical professionals to practice surgeries in risk-free environments, thereby improving patient safety and procedural outcomes [42]. Similarly, military operations leverage high-fidelity AR simulations to train personnel in battlefield environments, where real-time responsiveness and environmental mapping are crucial for mission planning and execution. Industrial sectors, including manufacturing, automotive design, and heavy machinery operation, also stand to gain from advancements in SLAM-based AR. In these environments, immersive simulators equipped with SLAM technology enable workers to engage with virtual prototypes, conduct safety drills, and troubleshoot machinery without physical risk. Automotive companies are increasingly integrating AR systems into driver training and vehicle testing, providing realistic driving simulations that reflect actual road conditions and vehicle dynamics [43], [44]. Enhancing the fidelity of these simulations through SLAM tracking improves operational safety, reduces errors, and enhances overall efficiency. A significant contribution of this research is its exploration of collaborative AR and VR workspaces. As industries shift towards remote collaboration and telepresence, synchronising multiple users within shared virtual environments becomes essential. This research lays the groundwork for advanced telepresence solutions, virtual meetings, and remote training programs by refining SLAM algorithms to accommodate multi-user interactions. Design, architecture, and engineering applications can benefit from synchronised AR environments, where participants can manipulate virtual prototypes in real-time [45].

Furthermore, this research contributes to the evolution of AR as a critical tool in future aviation practices. As the aviation industry progresses towards autonomous and semi-autonomous aircraft, SLAM-enhanced AR systems can support pilots in monitoring and managing increasingly automated cockpits. By integrating real-time environmental mapping with AR overlays, SLAM technology enables pilots to maintain situational awareness even in highly automated systems, bridging the gap between human oversight and machine precision. This capability is vital for successfully transitioning towards next-generation air traffic management systems, unmanned aerial vehicles (UAVs), and urban air mobility solutions. Ultimately, this research enhances the effectiveness of SLAM-based AR technology in aviation while laying a foundation for broader applications across multiple sectors. This study adds to the growing knowledge of augmented reality, virtual reality, and real-time localisation technologies by advancing SLAM algorithm development. These contributions are expected to drive future innovations, improving training fidelity, operational efficiency, and user experiences in dynamic, complex environments. The long-term impact of this research is anticipated to shape the future of AR applications, fostering safer, more immersive,

and practical training methodologies across aviation and beyond. The contributions of this study are summarized below:

- Comprehensive Review of SLAM Methods for AR Head Tracking: This study conducted an in-depth review of SLAM methods in augmented reality, emphasizing their potential to improve head-tracking accuracy in complex environments like flight simulators. Experiments were conducted to detect and extract features from video streams, with a focus on cockpit structures, addressing the limitations of current AR systems in achieving high spatial precision in enclosed and intricate environments.

- Evaluation of Marker-Based and SLAM-Based Tracking in Landing Scenarios: The study evaluated both marker-based and SLAM-based tracking methods within an AR environment to enhance training fidelity for landing maneuvers in pilot training. By testing these methods in scenarios critical to flight safety, this contribution helps bridge the gap in achieving high-precision tracking in complex and time-sensitive flight scenarios, where real-time accuracy is paramount for simulating realistic landing conditions.

- Adaptation of Fitts' Law to Assess Interaction Difficulty in AR Environments: To improve user interaction within AR-based simulators, this research adapted Fitts' Law to evaluate interaction difficulty across four quadrants of the AR environment. This contribution addresses cognitive load and usability challenges by optimizing the design of intuitive, responsive AR interfaces for flight training. Such adaptations ensure that AR systems in flight training environments align with pilots' cognitive and physical interaction requirements, thereby enhancing training efficacy.

- Head-Tracking Experiments with SLAM and Filter Evaluation: The study performed head-tracking experiments using SLAM algorithms within the Hololens 2 platform, focusing on four filtering techniques (Extended Kalman Filter, Sequential Importance Resampling, Unscented Kalman Filter, and Unscented Particle Filter) as well as fusion approaches. These experiments identified the most stable and reliable tracking solutions, which are crucial for maintaining accuracy and immersion in dynamic flight simulation environments. The findings are instrumental in advancing AR head-tracking technology for aviation simulators.

- Multimodal Biometric User Identification in AR Environments: The study developed a novel deep learning-based behavioural biometric system to identify users during AR interactions using gaze, head motion, and hand gesture data. A custom LSTM-based architecture processes temporal signals from the HoloLens 2 sensors to learn user-specific interaction patterns, enabling continuous, implicit, and privacy-preserving identification. This contribution supports seamless personalisation and user continuity within AR training environments, highlighting the role of AI in developing secure and adaptive AR systems.

- AI-Driven Behavioral Modeling and Cognitive State Prediction: The study developed a deep learning framework for predicting memory retention in AR

environments using behavioral data captured from Microsoft HoloLens 2 sensors. The framework analyzes key interaction metrics—gaze duration, interaction frequency, revisit counts, and head movement stability—to estimate users' memory states across four categories: Strong Recall, Weak Recall, Cognitive Overload, and Lack of Engagement. By training a neural network with category embeddings and optimizing its parameters through Bayesian tuning, the system achieved high prediction accuracy, aligning closely with actual user recall. This non-invasive, scalable model offers a significant advancement in real-time cognitive modeling, enabling AR systems to dynamically adapt content to users' engagement levels and memory retention likelihood. The findings have broad implications for personalized learning in AR-based flight training, medical education, and other simulation environments.

- AR-Enhanced Quantum Learning and Spatial Reasoning: The study extended AR's application to STEM education by developing an interactive simulation of the Mach-Zehnder Interferometer (MZI). Participants demonstrated a 45.45% improvement in conceptual understanding and a 70% success rate in reconstructing the MZI setup. The AR platform effectively visualized abstract quantum concepts such as superposition and interference, reducing cognitive load and enhancing engagement. This contribution showcases AR's broader potential beyond aviation, offering an accessible and intuitive tool for complex scientific learning.

1.4 AR Goggle (Microsoft HoloLens 2): Core Components and Functionality

The Microsoft HoloLens 2 is a stand-alone mixed-reality headset that overlays high-resolution holographic imagery onto the user's view of the real world. It is widely used in industrial, medical, and training contexts (e.g., surgical planning and flight simulation) because it can display interactive 3D content in situ. The device's advanced hardware – including see-through waveguide optics, multiple cameras and sensors, and a dedicated Holographic Processing Unit (HPU 2.0) – works together to align virtual content with physical surroundings. For example, HoloLens 2 supports fully articulated two-handed gesture tracking, eye-gaze tracking, and a field of view nearly double that of HoloLens 1. These upgrades enable users (such as pilots in a simulator) to intuitively interact with holographic instruments and maintain situational awareness during training [46].

Optical Display and Field of View (FOV): The optical system of the HoloLens 2 is a critical factor in achieving high levels of immersion and interaction fidelity. The device features industry-leading see-through holographic lenses, delivering 2K resolution per eye (2048 x 1080 pixels), which provides crisp and detailed virtual imagery essential for cockpit simulations. In aviation training, where pilots must interact with holographic representations of flight instruments and heads-up displays (HUDs), visual clarity directly impacts the accuracy of virtual overlays.

A broader field of view (FOV) is crucial in minimizing the cognitive dissonance between physical and virtual interfaces. The HoloLens 2 expands its FOV to 52

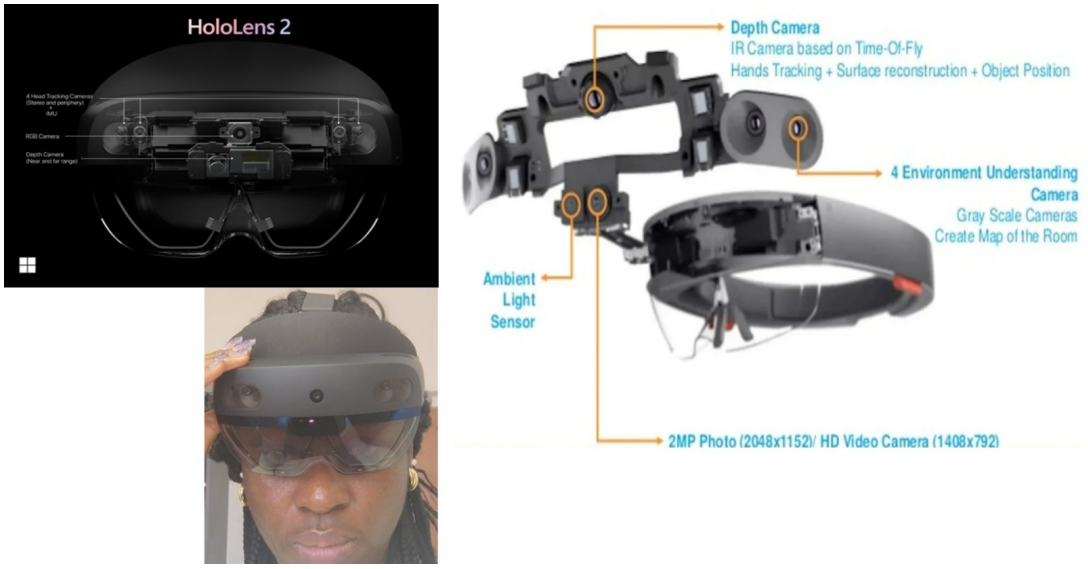


FIGURE 1.4: Detailed breakdown of the Microsoft HoloLens 2's core components, including head-tracking cameras, depth sensors, RGB camera, ambient light sensors, and environmental understanding cameras used for spatial mapping and object tracking.

degrees, compared to the 34-degree FOV of the original HoloLens 1, allowing for a more encompassing view of the cockpit environment. This enhancement reduces the need for frequent head repositioning, improving focus retention and user comfort during extended training sessions [47].

Head-Tracking and Spatial Awareness Sensors: HoloLens 2 achieves head tracking and spatial anchoring using an array of outward-facing sensors. The front visor houses four wide-FOV visible-light cameras for environment mapping and head motion sensing, one depth camera (time-of-flight) for range sensing, and one RGB color camera. Together these provide real-time visual-inertial SLAM (simultaneous localization and mapping). The device continuously scans the surroundings to build a 3D spatial mesh. This means virtual objects (e.g. instrument holograms) stay "fixed" to world positions even as the user turns or moves. In flight-simulation use, this keeps a holographic altimeter or heads-up symbology aligned with the cockpit even during rapid head movements. HoloLens 2 also includes a 5-channel microphone array and speakers for spatial audio (so virtual sound sources are anchored in space) [48].

Inertial Measurement Units (IMUs): In addition to cameras, HoloLens 2 uses an onboard IMU (Inertial Measurement Unit) comprising an accelerometer, gyroscope, and magnetometer. These sensors measure linear acceleration, angular velocity, and magnetic orientation, respectively. Fusing IMU data with the camera-based SLAM allows the system to maintain robust tracking even if visual features are few (e.g. in dark cockpits) or during very fast motions. In practice, if a pilot quickly looks down at a console or dives through clouds, the IMU fills in head-pose updates between visual frames, preventing drift of the holograms. The combination of optical SLAM and IMU thus provides 6-degree-of-freedom (6DoF) tracking, meaning the user's position and orientation are continuously tracked and virtual objects can remain anchored

accurately in 3D space [49].

Accelerometers – measures linear acceleration (x/y/z axes) and gravity.

Gyroscopes—measure angular rotation rates around each axis.

Magnetometers—provides an absolute orientation reference relative to Earth's magnetic field.

This sensor fusion ensures very low latency head tracking. Even with sudden head turns or in low light, the device can reliably track motion and keep holograms in place.

Optical Sensors and Environmental Mapping: The depth sensor and RGB camera also support environmental mapping. In "long-throw" mode, the time-of-flight depth camera scans the cockpit at low frame rates (1–5 fps) to construct a 3D point cloud of surrounding surfaces. This mesh is updated as the user moves, allowing holographic instruments to be pinned to physical panel surfaces. For example, an AR flight display can remain flush with the cockpit windshield or console even if the pilot shifts position. Because the sensors utilise active IR illumination, mapping is possible regardless of ambient light, thereby preserving the accuracy of hologram placement (drift in the virtual overlay is minimal). Essentially, the HoloLens 2 continually registers the real cockpit geometry, which is crucial for mixed reality: any misalignment between real and virtual elements would break immersion or confuse the pilot. By aligning digital flight aids with actual cockpit cues, situational awareness is enhanced and training scenarios remain realistic [49].

Eye and Hand Tracking: Natural Interaction in Simulated Cockpits: A key innovation in HoloLens 2 is built-in eye- and hand-tracking for intuitive control. The headset has two IR-based eye-tracking cameras (one per eye) which monitor the user's gaze in real time. With this, the user can select or activate virtual controls simply by looking at them (gaze can focus a reticle or initiate actions). In a cockpit simulation, a pilot might "dial" a holographic knob or read a procedural checklist by gaze alone. Hand tracking on HoloLens 2 is fully articulated (using 25+ tracked points per hand), allowing the user to gesture naturally. For instance, the pilot can grab or press a virtual switch, or swipe through instrument menus in mid-air. This direct manipulation of holograms – without any physical controller – is enabled by the same front cameras and depth sensor, which detect and interpret hand movements. (The original HoloLens supported only limited one-handed gestures, but HoloLens 2 supports bimanual interaction.) Together, gaze and gesture tracking make the interface far more fluid; a pilot no longer needs a separate keyboard or pointer to interact with AR cockpit overlays. All of these capabilities (hand, eye, and voice) are processed on-device in real time by a dedicated Holographic Processing Unit (HPU) to avoid lag [50].

Simultaneous Localization and Mapping (SLAM): HoloLens 2 continuously runs SLAM to lock virtual content to the world. As noted, the 4 tracking cameras provide visual-inertial SLAM by matching features frame-to-frame and fusing IMU data. This ensures that when a pilot looks around the cockpit, holographic indicators remain anchored to specific instruments or controls. In practice, SLAM creates a persistent spatial map ("spatial anchor") so that virtual overlays – such as navigation aids or emergency checklists – stay in the correct spot. For example, a holographic guide arrow fixed to the throttle quadrant will not drift away if the pilot leans forward. This robust world-anchoring is why HoloLens 2 is helpful for precise training: the system

"knows" the user's pose and environment well enough to consistently superimpose accurate visuals [48].

Processing Power and Connectivity: HoloLens 2 is powered by a Qualcomm Snapdragon 850 CPU coupled with a custom second-generation Holographic Processing Unit (HPU 2.0). The HPU handles all sensor processing (hand/eye tracking, spatial mapping, etc.) at very low power, while the main CPU runs applications. This combination enables the creation of complex mixed reality scenes without noticeable lag. The device features 4GB RAM, 64GB storage, Wi-Fi 802.11ac (Wi-Fi 5), and Bluetooth 5.0, enabling real-time network connectivity. For example, multiple trainees wearing HoloLens 2 can share the same virtual cockpit scenario via Azure cloud services or peer-to-peer, enabling collaborative training. The USB-C port and on-board battery provide mobility, though with only about 2–3 hours of active use, continuous power (or breaks) is required for longer sessions [51].

Environmental and Spatial Understanding: The HoloLens 2 provides robust six-degree-of-freedom (6DoF) inside-out tracking and real-time spatial mapping of the environment. Using multiple cameras, an inertial measurement unit (IMU) and a time-of-flight depth sensor, it continuously localizes itself in world space and scans surrounding surfaces. The result is a live 3D mesh of the physical. This mesh is used to anchor digital content to real objects: holograms stay fixed to walls or equipment, can be occluded by real surfaces, and even participate in physics simulations based purely on the captured geometry. [52]. In practice, HoloLens 2 tracks position and orientation (6DoF) while building a detailed mesh, so virtual and physical elements remain precisely aligned as the user moves. The device also supports:

Mixed Reality Capture: which enables the recording of photos or videos of the combined holographic and real-world scene, allowing for live documentation of training exercises or MR demonstrations.

6DoF Tracking: HoloLens 2 uses inside-out SLAM tracking to provide full six degrees of freedom. The headset's cameras and IMU continuously track head position and orientation on a global scale.

Real-time Spatial Mesh: Its depth sensor and cameras scan nearby surfaces to generate a high-resolution 3D mesh of the room. This mesh lets the system place and occlude holograms on real surfaces; for example, holograms automatically hide behind physical walls or objects based on the mapped geometry.

High-Accuracy Mapping: Studies show HoloLens 2 can reconstruct large indoor spaces with centimetre-level accuracy. For instance, Navares-Vázquez et al [53]. report that HoloLens 2 scans over 100 m trajectories with errors under 4 cm, generating point clouds suitable for building information models.

This tightly integrated tracking and mapping means the digital content "knows" the physical world. Spatial mapping has been shown to reliably capture room-scale geometry, enabling the precise alignment of virtual models and even contactless measurement of real dimensions. In effect, HoloLens 2 creates a digital twin of the environment in real-time. All holograms can then be anchored to this twin; for example, virtual instructions can be placed on real machines or equipment and will remain in place as users walk around. Mixed Reality Capture (the onboard camera) can then record these blended scenes: for example, capturing a trainee performing a task with overlaid holographic guidance.

Human Understanding and Interface: The device's voice command capabilities, combined with spatial audio and a five-channel microphone array, facilitate hands-free interaction and spatially aware sound experiences. This functionality enhances immersion in aviation training, enabling pilots to issue commands or interact with virtual elements without manually adjusting controls, supporting more intuitive and uninterrupted training sessions.

The advanced tracking systems, real-time spatial mapping, and robust processing power make the HoloLens 2 a powerful AR platform for flight simulation, industrial applications, and training scenarios that demand precision and responsiveness. While its capabilities provide substantial benefits for training environments, the device's limitations, such as battery life and potential sensor drift, must also be considered to assess its effectiveness fully in high-stakes environments like aviation.

1.4.1 Conclusion

This chapter has established the conceptual and technological foundation for integrating Simultaneous Localization and Mapping (SLAM)-based augmented reality (AR) goggles into flight simulation environments. By outlining the motivation, objectives, and methodological approach of the investigation, it sets the stage for a systematic investigation of how immersive AR systems can enhance training realism, situational awareness, and spatial cognition in aviation contexts. The detailed analysis of the Microsoft HoloLens 2 further underscores its suitability for real-time head tracking and interactive simulation tasks in high-fidelity training scenarios.

The next chapters provide a critical review of the existing literature on head-tracking technologies, SLAM algorithm development, and AR applications within aviation and related domains. The review identifies relevant theoretical frameworks, identifies current research gaps, and positions the present study within the wider landscape of immersive training technologies.

2 A review on tracking head movement in augmented reality systems

2.1 Introduction

Augmented reality (AR) represents a significant technological development that enables users to experience real-world environments enhanced with digital overlays. This emerging technology is rooted in advances in computer vision, sensor fusion, and spatial computing, all of which converge to facilitate the seamless integration of virtual content into physical spaces. By overlaying interactive digital elements onto the user's immediate environment, AR enriches the perceptual experience, fostering a hybrid reality where virtual objects coexist with the tangible world. These digital augmentations provide spatially consistent information about the observed scene by superimposing graphical objects aligned with real-world three-dimensional (3D) structures, as illustrated in Figure 2.1. This interaction between the digital and physical worlds enhances user experience and operational efficiency across various applications.

A critical AR capability is the real-time estimation of the camera's 3D position and orientation—often referred to as "pose"—relative to objects in the real world, which is derived from sophisticated computer vision techniques [55]. This process allows for precisely aligning virtual objects with the physical environment, enabling a seamless and immersive experience. The capacity to track and estimate pose in real-time has become a cornerstone of AR systems, directly influencing the accuracy and fluidity of virtual object placement. The adoption of AR has accelerated rapidly, driven by breakthroughs in hardware performance, computational power, and algorithmic advancements, thereby catalyzing innovation across diverse fields such as healthcare [56], education [57], tourism [58], design [59], and manufacturing [60]. In each of

[0]Publication note: This chapter is based on the peer-reviewed article [54]. https://doi.org/10.1016/j.procs.2023.10.431.

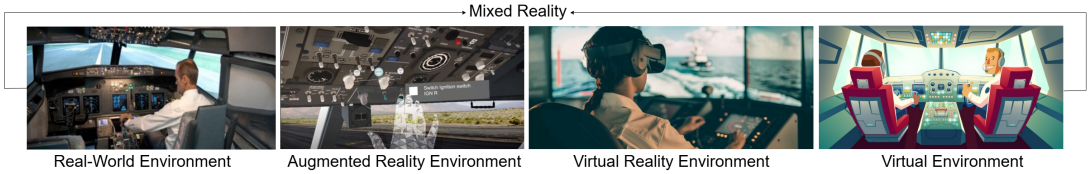


FIGURE 2.1: Reality-Virtuality Continuum.

these domains, AR's ability to superimpose virtual elements onto the real world with high precision has unlocked new possibilities, ranging from interactive learning environments and enhanced medical imaging to streamlined design processes and optimized production workflows. AR's unique ability to accurately overlay virtual elements onto the real world has revolutionized interactivity, training experiences, and operational efficiencies in various fields. As the technology has evolved to support real-time applications, its fusion of the virtual and physical realms has become particularly significant in contexts where situational awareness and precision are crucial, such as medical procedures, industrial design, and pilot training. As the technological landscape continues to evolve, AR is set to play an even more significant role in shaping the future of these industries, reinforcing its value as an indispensable tool for visualization, simulation, and collaborative interaction.

The potential of AR systems to accurately register and track virtual objects in dynamic environments is set to revolutionize user experiences with wearable devices. As AR technologies mature, the emphasis is increasingly on achieving seamless interaction through enhanced tracking and localization techniques. A fundamental component of this capability is head tracking, which allows AR devices to monitor the user's movements and position in the physical world, adjusting the virtual content accordingly to maintain precise alignment. This capability is pivotal in ensuring that virtual objects remain spatially consistent with the user's viewpoint, creating a cohesive and convincing augmented experience. Head tracking in AR relies on the continuous input and interpretation of real-time data regarding the user's physical location and the device's movement. This data stream is processed using advanced sensor arrays, including accelerometers, gyroscopes, and depth cameras, to generate accurate positional estimates. This process enables the precise placement and alignment of virtual objects within the user's field of view, achieving a seamless integration of virtual and real-world elements. The capacity to track head movements with low latency and high accuracy is essential for maintaining immersion, particularly in high-stakes applications such as military training simulations, remote surgical procedures, and complex manufacturing tasks. In practical applications such as training simulators or remote assistance, this integration ensures an immersive user experience and enhances situational awareness [46]. By minimizing discrepancies between the user's actions and the system's response, AR systems can foster deeper engagement and reduce cognitive dissonance, amplifying their effectiveness across many scenarios. Given the diverse applications of AR across industries, there is a need to overcome critical challenges to ensure accurate, real-time head tracking in diverse settings such as cockpit simulations, manufacturing floors, and healthcare environments. The complexity of dynamic environments introduces numerous challenges, including varying lighting conditions, occlusions, and the rapid movement of users, all of which can negatively impact tracking accuracy. Challenges like maintaining tracking accuracy under rapid head movements, occlusions, and changing environmental conditions impact the effectiveness of AR applications, where precise alignment is paramount. Tackling these challenges is essential for realizing the complete potential of AR across sectors. Various approaches, including predictive models, machine learning algorithms, and sensor fusion techniques, have been suggested to mitigate latency and improve tracking accuracy [61], [62]. These techniques leverage diverse data streams to enhance

tracking resilience, ensuring robust performance even in complex and unpredictable environments. However, the effectiveness of these techniques often varies based on particular use cases and environmental conditions.

This chapter aims to systematically review the latest head-tracking techniques in AR systems, evaluating their strengths and weaknesses across different real-world scenarios. By synthesizing current research, this work highlights areas where improvements are needed and identifies gaps in the literature to guide future research directions. This review establishes a foundation for the subsequent chapters in this thesis, which will explore new solutions to the identified challenges, advancing the development of more accurate, responsive, and resilient AR systems. The findings presented herein will contribute to the ongoing evolution of AR technology, positioning it as a transformative force in shaping the future of interactive and immersive user experiences.

2.1.1 Challenges and Advancement in AR Head Tracking

Head tracking is critical in augmented reality (AR) systems, facilitating the seamless interaction between users and virtual content in a manner that feels both natural and intuitive [63]. The precision of head tracking is instrumental in achieving the seamless integration of virtual objects with the physical environment, thereby enhancing the user's immersion and overall experiential quality. However, achieving accurate head tracking poses significant challenges, predominantly arising from human movement's dynamic and unpredictable nature. Head rotation, sudden directional shifts, and occlusions introduce significant variability in appearance, necessitating robust algorithms capable of maintaining spatial coherence and consistent content alignment across varying conditions. AR systems must effectively handle these variations by differentiating between distinct viewpoints—such as frontal facial perspectives and rear head orientations—while ensuring the preservation of accurate spatial mapping, even in the presence of significant environmental noise [64]. As such, achieving high accuracy, long-range tracking, and robustness in head tracking remains a critical area of research and development.

A principal challenge in AR head tracking is latency, the delay between initiating a user's head movement and the corresponding visual update presented within the head-mounted display (HMD). High latency disrupts the synchrony between user actions and virtual object responses, resulting in a disjointed experience that can induce discomfort and disorientation and negatively impact overall system usability [65]. Predictive modelling has emerged as a viable solution to mitigate this, offering the potential to forecast future head positions and preemptively adjust visual content. For instance, Gul et al. [61] proposed the application of Kalman filters for head motion prediction within a cloud-based volumetric video streaming framework, yielding prediction accuracies surpassing autoregressive models by 0.5 degrees over a 60-millisecond forecasting window. This innovative approach underscores the potential of predictive algorithms in minimizing perceived latency, enhancing user comfort, and bolstering AR system performance.

Recent research studies have pushed the boundaries of head tracking accuracy and robustness. Li et al. [62] introduced a comprehensive benchmark for evaluating

360-degree rotational head tracking, shedding light on the intricacies of real-world scenarios involving unrestrained, multidirectional head movement. Their employment of deep neural networks to refine head-tracking algorithms culminated in state-of-the-art accuracy, demonstrating the efficacy of machine learning-driven approaches in addressing complex tracking challenges. Furthermore, Lampropoulos et al. [66] integrated deep learning frameworks with semantic web technologies and knowledge graphs, creating AR systems that facilitate enhanced real-time interaction between users and digital content. This fusion of artificial intelligence and AR marks a significant leap forward in pursuing more responsive and intelligent augmented environments. To mitigate latency and improve tracking precision, hybrid tracking methodologies have garnered increasing attention within the AR research community. Researchers have explored hybrid frameworks that enhance adaptability across diverse AR applications by fusing Simultaneous Localization and Mapping (SLAM) techniques with inertial tracking systems. Syed et al. [67] demonstrated the efficacy of such hybrid systems in contexts spanning from medical imaging to military simulations, underscoring the versatility and robustness inherent in integrated tracking solutions. The synthesis of sensor data and visual cues provides a multifaceted tracking mechanism capable of maintaining accuracy across fluctuating environmental parameters. This ensures AR applications retain their integrity and functionality, even in challenging operational conditions. Machine learning algorithms have also proven invaluable in refining head-tracking performance. Attention-based models, in particular, have shown promise in predicting head movements with heightened accuracy, thereby reducing latency and enhancing real-time responsiveness. Lee et al. [68] explored the deployment of attention mechanisms to analyze and adapt to user behaviour patterns, resulting in smoother, more intuitive AR experiences. Similarly, Liu et al. [69] advanced head-tracking precision by leveraging DeepPose estimation models, which dynamically adjust to head positioning variations and compensate for real-time tracking discrepancies. These deep learning methodologies represent a paradigm shift in AR development, equipping systems with the capacity to evolve alongside user behaviour. Adaptive filtering techniques further exemplify the innovative strides made in AR head tracking. Ranjan et al. [70] devised dynamic filtering algorithms capable of recalibrating prediction models based on evolving user movement patterns, ensuring a consistently fluid AR experience. Complementarily, Chen et al. [71] investigated the application of computational offloading in edge computing environments to expedite data processing and minimize latency. Redistributing computational loads to peripheral devices optimizes system performance and facilitates the real-time synchronization of virtual content with user actions. Emerging research avenues continue to explore the potential of various deep learning frameworks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in augmenting pose estimation accuracy. Kumar et al. [72] conducted an extensive survey highlighting the effectiveness of these models in capturing intricate motion patterns and maintaining accuracy under suboptimal conditions, such as poor illumination and occlusion. Additionally, Jung et al. [73] pioneered multi-view stereo and real-time 3D reconstruction methodologies, leveraging stereo cameras and depth sensors to generate accurate 3D representations of user environments. This depth perception and spatial mapping integration enhances tracking fidelity and fosters more significant interaction between

users and virtual content.

These advancements address the dual challenges of latency and accuracy in AR systems, fostering more immersive, responsive, and reliable user experiences. As AR technology continues to mature, the proliferation of cutting-edge tracking solutions promises to expand its applicability across diverse sectors, ranging from healthcare and education to defence and entertainment.

2.2 Methods

This chapter employs a narrative literature review approach to examine recent academic and technical research on state-of-the-art head-tracking techniques and their applications in augmented reality (AR) systems. The review aims to systematically evaluate studies on head-tracking accuracy, latency reduction, real-time processing, and integration with emerging technologies such as machine learning and sensor fusion.

2.2.1 Analytical Framework

The review systematically identifies and analyzes relevant studies using a structured search and selection approach. Research articles were selected using specific inclusion criteria to ensure relevance to the study objectives. The focus is head-tracking methods that enhance accuracy, reduce latency, support real-time responsiveness, and employ innovative approaches like machine learning and sensor fusion. The search encompassed a variety of recent studies, prioritizing those published within the last decade to capture the latest advancements and empirical data on tracking performance in AR environments. The selected studies were categorized by head-tracking methodologies: vision-based, sensor-based, and hybrid approaches that integrate multiple tracking technologies. Within each category, subcategories were further developed to analyze distinct techniques, such as optical marker-based tracking, IMU-based sensor fusion, and SLAM-enabled tracking algorithms. This categorization provided a structured approach to assess each method's performance, including metrics such as mean tracking error, latency, frame rate, and spatial accuracy. Each category of tracking techniques was systematically evaluated to identify common trends, specific performance metrics, and challenges addressed, such as environmental noise, system latency, and computational demands. Where possible, comparative analysis techniques were applied to synthesize findings, allowing for a clear identification of strengths, limitations, and suitability for different AR applications. This structured analysis provides insights into existing gaps and potential areas for further development in head-tracking methods for AR.

2.2.2 Head Degrees of Freedom (DoF)

Degrees of freedom (DoF) in a tracking system represent the independent parameters that can be measured or tracked, determining the extent of movement within a system. In augmented reality (AR), DoF is crucial for head-mounted displays (HMDs), as it governs user interaction with virtual objects and significantly impacts immersion and

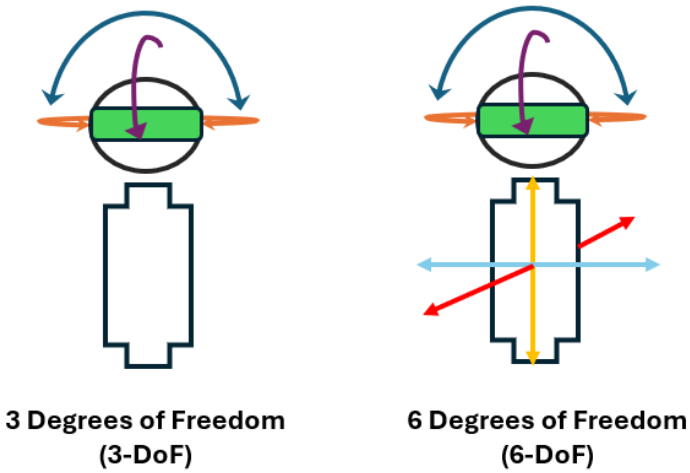


FIGURE 2.2: The figure shows 3 Degrees of Freedom (3DoF), which tracks only rotational movements (pitch, yaw, roll), with 6 Degrees of Freedom (6DoF), which tracks both rotational and translational movements (x, y, z axes), allowing for full movement and immersion in virtual environments.

realism. In 3D space, there are six degrees of freedom (6DoF), including rotational and translational movements [74]. Rotational movements—pitch, yaw, and roll—describe the orientation of the user's head. These are typically tracked by inertial measurement units (IMUs) such as accelerometers, gyroscopes, and magnetometers, which monitor changes in velocity and orientation to adjust the virtual environment as the user moves. The accurate tracking of these rotational changes is foundational to achieving spatial alignment in AR, as it ensures that virtual objects remain correctly oriented within the user's field of view [75]. Translational movements involve positional changes along three axes: forward/backwards (z-axis), left/right (x-axis), and up/down (y-axis), as illustrated in Figure 2.2. This type of motion tracking requires external cameras or built-in depth sensors, which are essential for positional tracking, allowing users to move freely within a 3D space [76]. Historically, AR and VR systems were limited to 3DoF tracking, which only detected rotational movements, keeping the user's position fixed. This limitation reduced immersion, as translational movements were not accounted for. However, modern 6DoF systems track rotational and translational movements, significantly enhancing the realism and interactivity of AR experiences [77].

Understanding and accurately tracking all six degrees of freedom are essential to achieving high levels of immersion in AR systems. In the context of this review, analyzing the technical aspects of 6DoF tracking enables a deeper evaluation of current tracking methodologies and their effectiveness in AR applications. IMUs, such as accelerometers, gyroscopes, and magnetometers, handle rotational movements, while depth sensors track positional changes. Understanding these components' technical and practical limitations is crucial for developing and optimizing AR systems to deliver an immersive, responsive user experience.

The systematic literature review conducted in this chapter forms the foundation for the subsequent chapters, which will explore innovative approaches to address

the identified challenges in head tracking. This review highlights opportunities for further research and development in AR systems by evaluating state-of-the-art tracking techniques, particularly improving head-tracking solutions' accuracy, latency, and responsiveness for real-world applications.

While this review provides a structured and comprehensive analysis, certain limitations must be acknowledged. The rapidly evolving nature of AR technology means that newer, potentially impactful studies may still need to be available in peer-reviewed journals. Additionally, the focus on empirical studies may exclude valuable insights from theoretical models that could inform future research directions. The findings presented here should be considered part of an ongoing research effort that requires continual updates as AR technology advances.

2.3 Tracking Techniques in Augmented Reality

This section explores the various tracking techniques employed in augmented reality (AR) applications, focusing on how AR systems utilize 6 Degrees of Freedom (6DoF) to track user movement and interaction within the environment. Accurate 6DoF tracking is essential for creating immersive AR experiences, as it allows for the precise alignment and stability of virtual objects within the physical space, enhancing user engagement and realism. Zhou et al. [76] categorize AR tracking methods into three main approaches: sensor-based, vision-based, and hybrid tracking methods, as shown in Figure 2.3.

Sensor-Based Tracking: Sensor-based tracking relies on embedded sensors such as accelerometers, gyroscopes, and magnetometers, which capture and monitor movement and orientation data in real-time [78]. This method effectively tracks rotational and translational movements, which are critical for applications requiring responsive and continuous tracking. However, sensor-based systems are often susceptible to cumulative error or drift, especially in long-duration sessions. Despite these limitations, sensor-based tracking plays a crucial role in monitoring user motion and is commonly combined with other tracking methods to improve accuracy.

Vision-Based Tracking: Vision-based tracking leverages camera systems to capture visual data, determining the user's or AR device's position and orientation relative to the physical world [79]. By processing image data, vision-based tracking provides visual feedback essential for object alignment and interaction within the AR environment. This method is particularly advantageous in settings where accurate spatial alignment of virtual objects is required. However, vision-based systems are highly dependent on environmental factors such as lighting conditions and are computationally intensive, which can result in latency or reduced performance in dynamic or poorly lit environments.

Hybrid Tracking Methods: Hybrid tracking methods integrate sensor-based and vision-based techniques to enhance tracking accuracy and robustness. By combining the continuous data stream from sensors with the spatial information from vision-based systems, hybrid tracking compensates for the individual limitations of each approach, resulting in more precise and reliable tracking solutions [80]. Hybrid methods are instrumental in applications where high precision and low latency are critical, as they

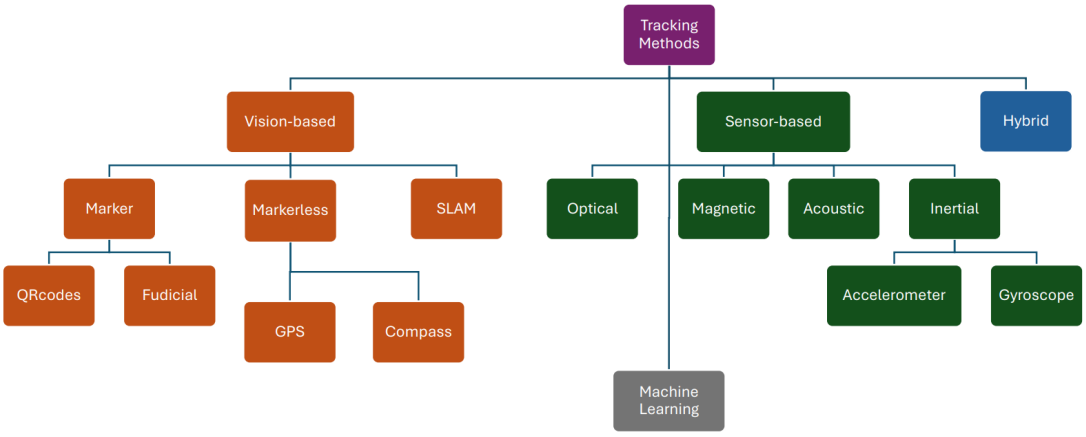


FIGURE 2.3: Taxonomy of Head tracking techniques in augmented reality

can adjust to environmental changes while reducing the effects of drift and minimizing latency.

The taxonomy in Figure 2.3 illustrates the three primary categories of head-tracking techniques in AR systems, providing a foundational understanding of their applications and limitations. These tracking methods have evolved to address the challenges of latency, spatial consistency, and environmental variability in AR, and each approach continues to be refined with technological advancements. This classification is a framework for analyzing current tracking systems and evaluating potential improvements to enhance AR experiences.

2.3.1 Marker-Based Tracking in Augmented Reality

Marker-based head tracking is a widely adopted technique in augmented reality (AR) systems, known for its reliability and precision, particularly in small or controlled environments. This approach leverages physical markers—typically fiducial markers or QR codes—placed within the environment, which the AR system's camera detects to provide a stable reference point. The AR system then uses computer vision algorithms to calculate the position and orientation of the user's head relative to the marker, enabling accurate and real-time tracking. Marker-based tracking is foundational in AR due to its straightforward implementation and consistent accuracy. It is popular in applications requiring precise spatial alignment, such as medical simulations and industrial training.

Techniques and Applications

Marker Detection: Marker detection involves identifying and interpreting physical markers through computer vision algorithms. Techniques like binarization and template matching are often employed to distinguish the marker from its background. Binarization simplifies the marker's image to black-and-white, enhancing contrast,

which aids in consistent detection across different lighting conditions. Template matching then compares the detected shape to predefined templates stored in the system, ensuring accurate identification. Advanced detection libraries, such as OpenCV's ArUco or AprilTag, enhance this process, allowing for robust recognition even in environments with partial occlusion or varying illumination [81]. Marker detection is crucial for ensuring the system's reliability and stability, as it forms the basis for subsequent tracking and pose estimation processes.

Tracking and Pose Estimation: After detecting the marker, the AR system calculates the user's head pose by determining the marker's orientation and position, facilitating six degrees of freedom (6DoF) tracking. This 6DoF capability allows the AR system to track rotational (pitch, yaw, and roll) and translational (up, down, left, right, forward, and backwards) movements, essential for accurately aligning virtual objects with the user's field of view. Precise pose estimation is critical in AR applications to avoid visual misalignments, which can disrupt the user experience. Some systems employ fiducial markers to resist occlusion and motion blur, providing reliable tracking even in complex environments where other methods might struggle [82]. This high level of accuracy in pose estimation is precious in medical applications and training environments where precise spatial alignment is paramount.

Rendering: Once the marker's position is tracked, the AR system overlays virtual content onto the user's display, rendering the augmented scene in real-time. This process is integral to achieving the immersive, interactive experience that AR applications aim to provide. Marker-based tracking excels in scenarios requiring consistent precision, particularly within controlled or structured spaces, such as interactive learning labs, medical training simulators, and industrial maintenance settings. In industrial contexts, markers can overlay virtual instructions or real-time data on equipment, providing technicians with contextual information without requiring them to divert their attention [79].

Applications and Limitations: Marker-based AR systems have demonstrated substantial utility in fields requiring high spatial accuracy. For instance, in medical settings, marker-based tracking enables the overlay of diagnostic information on a patient, assisting surgeons in real-time by maintaining spatial alignment of medical imagery during procedures [82]. In industrial training, markers augment machinery with instructional overlays, aiding technicians in maintenance tasks [81]. However, the dependency on physical markers poses limitations, especially in dynamic environments or larger spaces where installing markers may be impractical. Moreover, marker-based tracking may lack flexibility in settings with frequent or rapid changes in the field of view, requiring the marker to remain within the camera's sight.

Despite these limitations, recent advancements in AR research have begun addressing the constraints of marker-based tracking. For example, hybrid systems that combine marker-based tracking with other technologies, such as inertial sensors or Simultaneous Localization and Mapping (SLAM), have shown promise in improving adaptability and robustness in tracking across larger or more dynamic environments [37]. These advancements suggest that while marker-based tracking remains effective in specific applications, integrating it with additional tracking methods can enhance its versatility and usability in diverse AR applications.

FIGURE 2.4: The left image shows the fiducial marker used to align the virtual cockpit instrument, and the right image shows a screen playing a video of a landing manoeuvre over a customized marker.

2.3.2 Markerless-Based Tracking in Augmented Reality

Markerless tracking in AR eliminates the dependency on predefined physical markers, enabling systems to utilize natural environmental features or cues to estimate the position and orientation of the user's head. This advancement offers unparalleled flexibility, allowing AR systems to function effectively in diverse environments, including outdoor settings and dynamic spaces, where placing fiducial markers or QR codes is impractical. Markerless tracking leverages computer vision algorithms and auxiliary sensors, such as GPS, accelerometers, and gyroscopes, to provide real-time positional data. Depending on the algorithm used, markerless systems can deliver varying degrees of freedom (DOF), with advanced configurations supporting full 6-DOF tracking [83].

Techniques and Applications

Natural Feature Tracking (NFT): Markerless tracking often relies on Natural Feature Tracking (NFT), which involves detecting and tracking vital visual elements within the environment, such as edges, corners, or textures. These features, referred to as interest points, must exhibit robustness across different lighting conditions, viewpoints, and scales for reliable head tracking. Traditional algorithms such as SIFT (Scale-Invariant Feature Transform) and SURF (Speeded-Up Robust Features) have been widely used due to their ability to identify and describe features invariant to rotation and scale [84]. However, the computational overhead of these methods limits their application in mobile and resource-constrained AR systems. Recent advancements have introduced lightweight algorithms like ORB (Oriented FAST and Rotated BRIEF), which offer comparable performance at a fraction of the computational cost. Machine learning-enhanced feature detectors have also shown promise in adapting to complex environments by learning to prioritize salient features dynamically.

Simultaneous Localization and Mapping (SLAM) Integration: To enhance tracking accuracy, markerless AR systems often integrate with SLAM technologies. SLAM

simultaneously constructs a map of the environment while tracking the user's movement, enabling accurate head pose estimation even in large-scale or feature-scarce environments. Techniques like ORB-SLAM and Parallel Tracking and Mapping (PTAM) utilize image and depth data to continually update the user's position and orientation [85]. For instance, Huang et al. [86] proposed a stereo visual SLAM-based head tracking method that constructs sparse head maps using ORB feature extraction and stereo matching. This method calculates 3D-2D correspondences through projection matching, enhancing real-time performance and tracking accuracy. Madrigal et al. [87] presented a robust head pose estimation framework that integrates 2D and 3D cues. Their approach effectively handles partial occlusions and extreme head rotations, even in the presence of noisy depth data.

Depth Sensors and Object Recognition: Depth sensors, employing infrared or structured light, create 3D point clouds by measuring the distance between the device and objects in the environment. These sensors are particularly effective in environments with insufficient natural features for traditional NFT methods. For example, Cao et al. [88] proposed a system combining depth sensing and multimodal template matching to address the challenges of registering virtual objects in featureless environments. Despite its advantages, markerless tracking faces several limitations:

1. Computational Overhead: Real-time processing of the environment and user movement imposes significant demands on computational resources, particularly in mobile AR applications.

2. Environmental Sensitivity: Tracking accuracy can degrade under poor lighting conditions, uniform textures, or highly reflective surfaces.

3. Occlusion Handling: Partial or complete occlusion of critical features by the user's body or external objects remains a persistent challenge.

While markerless tracking systems offer flexibility and scalability, they face computational and environmental challenges. Future advancements could focus on reducing computational overhead by integrating lightweight algorithms or leveraging hardware accelerators. Moreover, combining depth sensors with advanced feature-detection methods may enhance robustness in low-texture or poorly lit environments. These adaptations would allow markerless tracking to operate more efficiently across a broader range of AR applications.

2.3.3 Optical Sensor Tracking

Optical sensor tracking is a three-dimensional (3D) localization system widely used in augmented reality (AR) and virtual reality (VR) applications to enable precise head movement tracking and seamless user interaction with digital environments. By leveraging cameras and sensors, this technology provides real-time positional and orientational data critical for delivering immersive experiences. A typical optical tracking system comprises a camera and a tracking sensor, both essential for capturing and analyzing user movements. The camera, often mounted on the head-mounted display (HMD) or mobile device, captures images of the surrounding environment. These images are processed by the tracking sensor using advanced algorithms, such

as the Kalman filter, which is renowned for its predictive accuracy. The Kalman filter refines its predictions of the camera's future position and orientation by incorporating prior data and accounting for noise and measurement errors, enhancing the system's reliability in real-time tracking scenarios [89].

Despite their effectiveness, optical sensor tracking systems face significant challenges in real-world applications. High computational demands and sensitivity to environmental factors, such as variable lighting and optical noise, can limit their performance. Occlusion, where objects or the user's body obstruct the camera's view, further exacerbates these limitations. Recent advancements have addressed these issues using structured light and infrared (IR) technologies. Structured light methods project patterns onto the environment and analyze their distortion to derive depth and spatial information, improving tracking precision [90]. Similarly, using IR pass filters and IR LEDs has been demonstrated to effectively mitigate the effects of ambient lighting by creating a controlled illumination environment [91].

Optical tracking systems typically support six degrees of freedom (6-DOF), encompassing three translational movements (x, y, and z axes) and three rotational movements (pitch, roll, and yaw). The degree of precision and freedom achievable depends on the system's configuration, with multi-camera setups providing higher accuracy and reliability.

While optical sensor tracking is highly effective in controlled environments, challenges such as latency, sensitivity to external conditions, and computational overhead remain active research areas. Hybrid systems combining optical tracking with other methods, such as inertial sensors or SLAM, are emerging as promising solutions for addressing these issues. Despite these challenges, optical sensor tracking is vital in advancing AR and VR technologies, enabling applications ranging from industrial training to healthcare simulations.

2.3.4 Magnetic Sensor Tracking

Magnetic tracking is a widely used technique for determining the position and orientation of an object within a specific environment through the detection of magnetic fields. This technology operates by placing a sensor on a target object, such as a user's head. At the same time, a set of coils or magnets generates a magnetic field in the surrounding environment [92]. As the user moves, the sensor captures variations in the magnetic field, transmitting this data to a computer for real-time processing. This capability accurately aligns virtual elements with the user's physical surroundings, enhancing the immersive experience in augmented reality (AR) environments.

In head orientation tracking, Wohle et al. [93] introduced a novel method that leverages eye-tracking data to enhance the accuracy of Magnetic Angular Rate Gravity (MARG) sensor-based orientation estimation. This approach is relevant in environments prone to magnetic field disturbances. Their system combines a MARG sensor data fusion filter, an online visual fixation detection algorithm, and a dynamic angular rate threshold estimation system. A notable advantage of their method is its self-contained design, which eliminates the need for stationary environmental references, allowing it to function autonomously. Benchmarking against the Qualisys motion capture system demonstrated superior heading accuracy, even under challenging conditions of magnetic interference. Singh et al. [94] proposed a three-dimensional

position and orientation tracking system that integrates a single-axis magnetic field transmitting inductor coil, a tri-axis receiving inductor coil, inertial measurement units (IMUs), and microcontroller units (MCUs). This system achieved remarkable accuracy, maintaining position jitter below 1 mm at distances of up to 1 meter, making it highly suitable for applications requiring precise positional data. Similarly, Kataria et al. [95] developed a six-degrees-of-freedom (6-DOF) magnetic motion tracking device that incorporates a self-healing neural model (SHNM). This model compensates for data loss caused by magnetic field interference, enabling seamless real-time tracking of a pilot's head and ensuring flawless synchronization with helmet-mounted displays (HMDs). Such innovations highlight the versatility and potential of magnetic tracking in high-stakes, real-time AR applications.

Magnetic tracking offers several key advantages. Its high responsiveness and accuracy, particularly in real-time applications, make it a reliable choice for systems requiring precise and dynamic positional data. Furthermore, magnetic tracking systems are relatively compact and consume less power than other methods, such as optical tracking, making them ideal for small, portable devices. However, the technology has limitations. Magnetic tracking is particularly susceptible to jitter as the distance between the sensor and the magnetic field source increases. It is also vulnerable to electromagnetic interference, which can significantly degrade accuracy in environments with high levels of electronic noise. Additionally, the degrees of freedom (DOF) achievable depend on the number and configuration of magnetic sensors within the system, with most systems offering 6-DOF similar to optical tracking [74].

Despite its challenges, magnetic tracking remains a cornerstone in developing advanced AR systems, particularly in applications where precise, real-time tracking is critical. Future research may address the limitations of electromagnetic noise and jitter through improved sensor designs and algorithmic innovations, such as advanced filtering techniques or hybrid approaches that combine magnetic tracking with complementary systems like inertial or optical sensors. These advancements could enhance magnetic tracking's utility across diverse AR applications, including navigation, training simulations, and interactive entertainment.

2.3.5 Acoustic Sensor Tracking

Acoustic tracking leverages sound waves to determine the position and movement of objects in physical space and has emerged as a valuable method in augmented reality (AR) systems. By precisely locating virtual objects within the real-world environment, this technology enhances the immersive quality of AR experiences. Accurate alignment of virtual and physical elements—an essential factor for user interaction and perception—is significantly facilitated by acoustic tracking [96].

One of the most widely adopted approaches to acoustic tracking involves ultrasonic sound waves. In this method, ultrasonic emitters are strategically placed within the physical environment, and the AR device, such as a smartphone or a head-mounted display (HMD), is equipped with sensors to detect these signals. By measuring the time-of-flight (ToF)—the duration taken by sound waves to travel from the emitters to the sensors—the system calculates the user's position and orientation relative to the emitters. This data is then used to accurately position virtual objects in the AR display, ensuring their alignment with the physical environment [97]. An alternative approach

employs the AR device's built-in microphones to capture environmental sounds. For example, a speaker emitting a specific tone or frequency can be the sound source. The AR device uses signal properties, such as time delay and frequency shifts, to estimate its position relative to the source. While this method can provide accurate tracking, it is generally slower than other sensor-based techniques and is more susceptible to environmental factors like ambient noise and physical obstructions.

Recent research has yielded innovative acoustic tracking systems that address some of the limitations of earlier methods, broadening the potential applications of this technology. For instance, Ge et al. [98] proposed an Acoustic Strength-based Angle Tracking (ASAT) system, which uses the strength of received acoustic signals to estimate the angle of arrival and subsequently determine the sound source's location. Their system achieved localization accuracy within 5 cm at distances of up to 3 meters, highlighting its potential for low-cost, accurate motion tracking in AR applications. Similarly, Wang et al. [99] introduced MilliSonic, a system capable of simultaneously tracking up to four smartphones without compromising accuracy or frame rate. By enabling 3D tracking for platforms like Google Cardboard, MilliSonic facilitates more complex interactions and expands the scope of AR applications. These advancements underscore the growing importance of acoustic tracking in developing cost-effective, scalable AR systems. Despite its strengths, acoustic tracking has its challenges. Environmental elements, including changes in temperature and humidity, ambient noise, reverberation, and physical obstructions, can adversely impact tracking accuracy. For example, changes in airspeed due to temperature or humidity fluctuations can alter the time-of-flight and frequency shift measurements, introducing errors in position estimation. Additionally, acoustic tracking systems typically offer either three degrees of freedom (3-DOF) or six degrees of freedom (6-DOF), depending on the configuration and number of acoustic sensors used [100]. While 6-DOF systems provide comprehensive tracking capabilities by capturing both positional and orientational data, they may require more complex setups, increasing system cost and computational demands.

Acoustic tracking holds considerable promise for AR applications requiring cost-effective and scalable solutions. Future advancements could address current limitations by integrating advanced noise-cancellation algorithms, robust signal processing techniques, and hybrid tracking methods that combine acoustic data with inputs from inertial measurement units (IMUs) or vision-based systems. These enhancements have the potential to improve tracking accuracy, expand degrees of freedom, and mitigate environmental sensitivities, further cementing the role of acoustic tracking in the evolution of AR technologies.

2.3.6 Inertial Sensor Tracking

Inertial tracking is a method for monitoring an object's movement and orientation using sensors that measure acceleration, velocity, and angular velocity changes. Widely applied in robotics, virtual reality (VR), and augmented reality (AR), inertial tracking provides precise motion tracking in three-dimensional space, making it an indispensable tool in these fields. Inertial tracking systems typically employ a combination of accelerometers, gyroscopes, and magnetometers, which work together to provide comprehensive data on an object's motion and orientation. These sensors may be

integrated into a single unit or distributed across a device to enhance tracking accuracy [101].

Compared to optical tracking, which relies on external cameras or markers, inertial tracking offers several notable advantages. It is inherently mobile, as it does not depend on external references, and it can operate effectively in outdoor or low-light environments where optical systems might fail. Furthermore, inertial tracking systems are compact and well-suited for portable and wearable devices. However, inertial tracking is challenging. A fundamental limitation is drift, where minor measurement errors accumulate over time, leading to significant inaccuracies in the object's tracked position and orientation. This issue is particularly pronounced in long-term tracking applications. Drift occurs because inertial tracking relies on integration calculations to estimate position and orientation, amplifying sensor measurement errors. Consequently, systems that rely solely on inertial tracking require periodic recalibration or augmentation with complementary tracking methods to maintain accuracy.

To address these limitations, recent research has focused on innovations to improve the accuracy and reliability of inertial tracking systems. Muchallil et al. [102] proposed integrating inertial measurement units (IMUs) into head-mounted displays (HMDs) to reduce latency in AR applications. Their approach enhances user immersion by ensuring that virtual elements respond seamlessly to head movements in real time. Jambrošić et al. [103] developed a binaural synthesis system employing IMUs for head tracking in VR and AR environments. This system dynamically calculates binaural signals based on head orientation, significantly improving the spatial audio experience. Similarly, Jadid et al. [104] explored using multiple low-cost IMUs to enhance tracking accuracy in mobile mixed reality systems. By fusing data from IMUs arranged on planar and non-planar grids, their system achieved tracking precision comparable to high-cost IMUs, demonstrating the potential of multi-sensor setups for cost-effective, accurate inertial tracking.

Inertial tracking systems typically provide six degrees of freedom (6-DOF), capturing both rotational (pitch, roll, and yaw) and translational (x, y, and z) movements [105]. The degrees of freedom and overall accuracy depend on the inertial sensors' number, quality, and configuration. Systems with well-calibrated sensors and advanced fusion algorithms can achieve high precision even in challenging environments. Despite challenges like drift, inertial tracking remains vital for applications requiring precise, real-time movement tracking in AR and VR. Its portability and ability to function without external references make it particularly valuable for scenarios where other tracking methods are impractical, such as outdoor navigation, mobile AR games, and industrial training simulations.

Ongoing research seeks to enhance inertial tracking by integrating complementary systems to mitigate drift and improve robustness. For instance, hybrid tracking solutions combining inertial data with optical or SLAM-based systems have shown promise in maintaining long-term accuracy. Additionally, applying machine learning algorithms to sensor data processing offers the potential for adaptive error correction and improved tracking reliability. These advancements position inertial tracking as a cornerstone technology in the evolving landscape of AR and VR.

2.3.7 Machine Learning in Augmented Reality Tracking

The precision and robustness of head tracking in augmented reality (AR) can be significantly enhanced by applying machine learning algorithms. These algorithms predict head movements based on user actions and effectively filter out noise and erroneous data from sensors such as accelerometers, gyroscopes, and cameras. Machine learning-based techniques offer greater adaptability than traditional tracking methods, which often struggle in low-light environments or with occlusions and sudden movements. By training models on large datasets of annotated images, these algorithms can recognize and track facial features, such as eye positions or nose direction, to ensure accurate head tracking even in challenging conditions.

Machine learning dynamically enhances real-time head-tracking performance by adapting processing settings and algorithms to user behaviour. This adaptability enables AR systems to respond more effectively to rapid or complex movements. Deep learning and convolutional neural networks (CNNs) are among the most commonly used machine learning techniques for AR tracking, as they allow accurate and efficient estimation of head position and orientation. For example, Yoon et al. [106] developed a 6 degrees of freedom (DOF) pose prediction algorithm designed for lightweight AR glasses. Their approach combined a consistent motion model with deep learning techniques, including controlled residual learning and model ensemble, to mitigate 3D visual inconsistencies caused by remote rendering latency. The algorithm achieved precise predictions in under 200 milliseconds, demonstrating its potential for real-time AR applications. Similarly, Kastner et al. [107] proposed a deep learning-based calibration method for AR devices that utilized data from 3D depth sensors. Xu et al. [108] introduced a novel technique for head posture estimation using deep neural networks and 3D point clouds generated from depth sensors. Their method divided 3D head pose angles into 36 classes, with 5-degree intervals, and employed a multi-layer perceptron to predict the probability of each angle. Using a graph convolutional neural network, their system achieved computational efficiency and reduced latency, distinguishing it from traditional methods reliant on 2D RGB or depth images.

The degrees of freedom (DOF) in machine learning-based tracking systems for AR depends on the algorithm used and the quality of input data. Systems leveraging deep learning often achieve full 6-DOF tracking by capturing rotational (pitch, roll, and yaw) and translational (x, y, and z) movements. However, the effectiveness of these systems is influenced by several factors:

1. Scene Complexity: Machine learning models must handle diverse and dynamic environments, which can introduce variability in tracking performance.

2. Noise and Occlusion: High levels of noise or occlusion in the sensor data require robust filtering and predictive capabilities.

3. Model Efficiency: The computational efficiency of the model determines its suitability for real-time applications, particularly on resource-constrained devices.

While machine learning significantly advances AR tracking, several challenges remain. Training robust models requires large, diverse datasets, and ensuring real-time performance on lightweight devices is computationally demanding. Future research could explore:

Hybrid Models: Combining machine learning with traditional tracking techniques, such as SLAM or inertial tracking, to reduce drift and improve robustness.

Edge Computing: Deploying lightweight machine learning models optimized for edge devices to enable real-time processing without reliance on cloud-based systems.

Adaptive Algorithms: Developing algorithms that can self-optimize based on environmental changes, improving performance in variable conditions.

Machine learning continues to be a transformative technology in AR, addressing challenges that traditional methods struggle to overcome. Advancements in real-time tracking capabilities and improvements in accuracy under diverse conditions lay the groundwork for more immersive and reliable AR experiences.

2.4 Results and Discussion

As outlined in the introduction, augmented reality (AR) has emerged as a transformative tool across diverse fields, demanding accurate head tracking for seamless alignment of virtual and real-world elements. This study examines various AR tracking techniques, highlighting their advantages, limitations, and suitability for specific applications. A comparative summary is provided in Table 2.1, illustrating the trade-offs between accuracy, speed, sensitivity, degrees of freedom (DOF), and environmental adaptability.

Key Insights from Comparative Analysis

1. Vision-Based Tracking: Marker-based systems are ideal for precision tasks in controlled environments due to their high accuracy and efficiency. However, their reliance on predefined markers limits scalability for larger or dynamic spaces.

   Markerless tracking offers enhanced flexibility by leveraging natural features, making it suitable for complex and large-scale environments. However, its computational intensity can challenge real-time applications on resource- constrained devices.

   SLAM techniques stand out as the most adaptable vision-based methods, performing well in dynamic and variable environments. Despite their strengths, they require significant computational power, necessitating optimization for mobile platforms.

2. Sensor-Based Tracking:

   Optical tracking excels in accuracy and dependability but suffers from occlusion and optical noise, reducing its applicability in unstructured settings.

   Magnetic and acoustic methods offer cost-effective solutions for specific use cases but are hindered by sensitivity to environmental factors like electromagnetic interference and humidity.

   Inertial tracking provides robust portability and independence from external references. However, drift remains a critical limitation for long-term applications, highlighting the need for periodic recalibration or hybrid approaches.

TABLE 2.1: Concise Comparative Analysis of AR Tracking Techniques

Technique	**Accuracy**	**Speed**	**Sensitivity**	**DOF**	**Environment**
Marker-based	High	Fast	Marker visibility, camera resolution	6DOF	Controlled
Markerless	Moderate	Slow	Motion, lighting conditions, surface texture	6DOF	Complex
SLAM	High	Fast	Environmental complexity, sensor accuracy	6DOF	Dynamic
Optical	High	Slow	Optical noise, occlusion, lighting conditions	6DOF	Controlled
Magnetic	Moderate	Fast	Electromagnetic interference, nearby electronic devices	6DOF	Metallic
Acoustic	Moderate	Slow	Temperature, humidity, ambient noise	3/6DOF	Outdoor
Inertial	Moderate	Fast	Drift, noise, vibration	6DOF	Small Spaces, Various
Hybrid	High	Fast	Magnetic interference, lighting changes	6DOF	Complex
Machine Learning	High	Fast	Lighting, camera quality, occlusion, facial features	3/6DOF	Multi-platform, Various

3. Machine Learning-Based Tracking: Machine learning approaches offer unmatched adaptability and accuracy, excelling in scenarios with occlusions, dynamic conditions, and complex environments. Their reliance on extensive datasets and computational resources, however, underscores the need for advancements in hardware efficiency.

The suitability of a tracking method is highly dependent on the specific requirements of the AR application. While marker-based and optical systems are well-suited for precision tasks in controlled environments, hybrid approaches integrating SLAM, inertial tracking, and machine learning emerge as the most effective solutions for dynamic and large-scale applications. Future research should prioritize:

Optimizing hybrid systems to balance accuracy with computational efficiency. Developing energy-efficient algorithms and hardware to support machine learning-based tracking on mobile devices. Exploring innovative sensor fusion techniques to overcome current environmental sensitivities and performance limitations.

2.5 Conclusion

The development of augmented reality (AR) systems has driven significant advancements in precise and reliable head-tracking methods for head-mounted devices tailored to indoor and outdoor environments. This chapter presented a comprehensive comparative analysis of various head-tracking techniques, focusing on critical parameters influencing their performance, including accuracy, responsiveness, environmental adaptability, and computational requirements.

The review highlighted critical challenges, such as occlusion, environmental noise, lighting variability, and sensor limitations, which affect the effectiveness of head-tracking systems. Understanding these factors is pivotal for selecting and optimizing tracking methods suited to specific AR applications. Achieving optimal tracking performance demands a deep awareness of these challenges and a thorough knowledge of available sensors and methodologies and their trade-offs.

The insights derived from this analysis will directly inform the design and implementation of experimental work to improve head-tracking accuracy in AR systems. By identifying the strengths and limitations of existing techniques, this review lays the groundwork for developing an advanced algorithm that addresses critical challenges, including latency, drift, and jitter. Such an algorithm will be essential for achieving smooth, responsive, and reliable head tracking in dynamic and complex environments.

Future experiments will focus on designing a robust and adaptive tracking system capable of maintaining performance across varying lighting conditions and environmental complexities. The goal is to ensure immersive AR experiences by seamlessly aligning virtual objects with real-world perspectives, even under challenging circumstances. This research will advance AR technology, paving the way for broader adoption and enhanced functionality in education, healthcare, industrial training, and entertainment.

3 SLAM Methods for Augmented Reality Systems for Flight Simulator

3.1 Introduction

Simultaneous Localization and Mapping (SLAM) has long been a focal point of research in robotics and intelligent systems due to its critical role in enabling autonomous navigation within complex environments [13]. Researchers have been working to address the challenges associated with SLAM for over three decades, contributing to developing numerous algorithmic solutions. Although significant progress has been made, SLAM remains a computationally intensive problem, with approximations evolving to address its inherent complexity [109]. This problem, which has persisted for over 30 years, has seen various approximations come close to resolving its challenges following decades of mathematical and computational efforts . Resolving the SLAM problem has the potential to unlock a vast array of applications in autonomous systems, particularly in robotics, where accurate localization and mapping are indispensable for independent operation in dynamic and unpredictable settings [110]. A robot capable of navigating its environment without human intervention—autonomy—requires a thorough understanding of its surroundings and consistent, accurate positional tracking.

SLAM systems integrate diverse sensor inputs, including lidar, cameras, inertial measurement units (IMUs), odometry, and GNSS receivers, to construct a map of the environment and localize the device. Visual SLAM, which relies on camera sensors, and LiDAR-based SLAM, which utilizes laser scanners, have been among the most prominent approaches [111]. These systems have demonstrated significant utility across various domains, including autonomous driving and augmented reality (AR). Within AR, SLAM is pivotal in ensuring virtual objects seamlessly align with the physical environment, enhancing user immersion and interaction.

Furthermore, the solutions to the SLAM problem can be categorized as filter-based and global optimization approaches. Filter-based methods, such as Kalman filters (EKF, UKF, SEIF) and particle filters, predict motion using IMU data and refine it with sensor measurements [112]. Global optimization approaches, such as bundle adjustment, estimate robot trajectories with higher accuracy but greater computational demands [113]. Vision-based SLAM techniques like ORB-SLAM leverage these methods to estimate the camera pose and map the environment effectively [114].

An exemplary application of SLAM in AR is the Microsoft HoloLens, a head-mounted device that projects augmented content into the user's field of view. The

[0]Publication note: This chapter is based on the peer-reviewed article [16]. https://doi.org/10.1007/978-3-031-35995-8_46.

HoloLens employs a variant of SLAM that integrates accelerometers, gyroscopes, and magnetometers to provide accurate 3D mapping and real-time tracking capabilities [115]. This integration enables the stable placement of virtual objects, ensuring they remain anchored within the natural world as the user moves. SLAM on Microsoft HoloLens builds 3D maps of an environment while tracking the device's position in real time, employing RGB cameras, depth sensors, and IMUs. Advanced SLAM techniques like RGB-D SLAM and Visual Inertial SLAM (VISLAM) enhance its accuracy and stability. Moreover, Microsoft has focused on large-scale scene reconstruction with voxel hashing and RGB-D camera re-localization [116].

SLAM methodologies have evolved through numerous breakthroughs, including MonoSLAM [117], ORB-SLAM [118], LSD-SLAM [31], and Visual Inertial SLAM (VISLAM) [119]. These techniques have targeted various objectives, such as motion estimation and scale ambiguity. Deep learning has also emerged as a powerful tool for enhancing SLAM by improving feature extraction and adaptability to complex environments. In the context of augmented reality for flight simulators, SLAM presents unique opportunities and challenges. Flight simulation environments are characterized by complex and dynamic settings, necessitating high-precision tracking to ensure that virtual objects align accurately with the cockpit's real-world features. SLAM's ability to address static and dynamic registration issues makes it an invaluable tool for AR-based flight simulation [120]. By enabling immersive and stable AR experiences, SLAM enhances the fidelity of training simulations, supporting the creation of more engaging and effective educational resources. Studies like Skurowski et al. **27** have explored SLAM in flight simulators, utilizing markers such as QR codes to avoid adding non-natural elements to cockpit environments.

SLAM methodologies have evolved through numerous breakthroughs, including MonoSLAM [117], ORB-SLAM [118], LSD-SLAM [31], and Visual Inertial SLAM (VISLAM) [119]. These techniques have targeted various objectives, such as motion estimation and scale ambiguity. Deep learning has also emerged as a powerful tool for enhancing SLAM by improving feature extraction and adaptability to complex environments. In the context of augmented reality for flight simulators, SLAM presents unique opportunities and challenges. Flight simulation environments are characterized by complex and dynamic settings, necessitating high-precision tracking to ensure that virtual objects align accurately with the cockpit's real-world features. SLAM's ability to address static and dynamic registration issues makes it an invaluable tool for AR-based flight simulation [120]. By enabling immersive and stable AR experiences, SLAM enhances the fidelity of training simulations, supporting the creation of more engaging and effective educational resources. Studies like Skurowski et al. [121] have explored SLAM in flight simulators, utilizing markers such as QR codes to avoid adding non-natural elements to cockpit environments.

This study analyses several key methodologies, including RGB-D SLAM, ORB-SLAM, LSD-SLAM, Visual Inertial Navigation (VIN), and deep learning-based SLAM. It focuses on applying SLAM in AR flight simulators. These algorithms are assessed according to their computational efficiency, accuracy, and adaptability to challenging conditions such as varying lighting, occlusions, and cockpit complexity. RGB-D SLAM, in particular, is employed to extract environmental features from live-stream video captured by the Microsoft HoloLens 2, focusing on the cockpit's spatial characteristics.

The insights derived from this research have broader implications for developing SLAM-based AR systems. Future advancements should prioritize the optimization of SLAM algorithms to reduce latency and computational demands, enabling real-time performance on resource-constrained devices such as AR headsets [113]. Addressing the impact of environmental variables, such as lighting changes and occlusions, remains crucial for improving robustness and reliability. By exploring the role of SLAM in AR-based flight simulators, this study evaluates and compares Microsoft HoloLens SLAM techniques against other SLAM and deep learning methodologies. The results are intended to guide the creation of more stable and immersive AR experiences, ultimately supporting the broader application of SLAM in training, education, and simulation.

3.2 Materials and Methods

The primary objective of this research is to explore and evaluate the effectiveness of various SLAM methods in improving head-tracking accuracy within AR-based flight simulators. By integrating SLAM algorithms into AR systems, this study aims to enhance the fidelity of virtual object placement in cockpit environments, ensuring seamless interaction between the trainee pilot and the AR system. This section provides a detailed description of the SLAM algorithms used, the experimental setup, and the evaluation criteria adopted for this research.

3.2.1 SLAM Algorithms for AR in Flight Simulators

This study evaluates five SLAM algorithms to assess their effectiveness in augmented reality (AR) flight simulators: RGB-D SLAM, ORB-SLAM, LSD-SLAM, Visual-Inertial SLAM (VINS), and deep learning-based SLAM. These algorithms were chosen for their established reliability in tracking and mapping tasks and their capacity to tackle important issues related to head tracking in dynamic environments.

One of the selected SLAM algorithms was implemented on the Microsoft HoloLens 2, a state-of-the-art head-mounted device with advanced sensors. The HoloLens 2 provided real-time video streams and depth data from the cockpit environment, enabling a comprehensive evaluation of the algorithm's performance in extracting and tracking environmental features. The choice of this platform was strategic, as it simulates the real-world conditions of AR-based flight training, including challenges such as drift, latency, and spatial misalignment.

The selection of these SLAM methods reflects their practical relevance and their ability to address the demands of AR in flight simulators. Stable and accurate feature extraction is critical for precise mapping and localization within the cockpit, ensuring seamless integration of virtual objects with the physical environment. These capabilities are directly linked to the overall system performance and the quality of the immersive experience provided to the user. By evaluating these algorithms, this study contributes to advancing AR technologies for training and simulation, with potential implications for improving pilot training efficacy and system reliability.

3.2.2 RGB-D SLAM

RGB-D SLAM technology creates a live 3D representation of the environment by combining RGB (colour) and D (depth) data. By incorporating depth information, RGB-D SLAM overcomes challenges commonly encountered by conventional RGB-only SLAM methods, such as handling repetitive or texture-less surfaces. The system uses an RGB-D camera to capture the colour and depth of each pixel. It applies sophisticated algorithms to estimate the camera's position while simultaneously building a 3D representation of the environment as the camera moves [122].

Modern RGB-D SLAM systems align spatial point features using depth data provided by sensors. This alignment is typically achieved with the Iterative Closest Point (ICP) algorithm, a widely utilized technique in computer vision that iteratively minimizes the differences between two point clouds [123]. The algorithm ensures coherence in the generated map by matching these points. However, a notable limitation of RGB-D SLAM is the noise inherent in in-depth measurements, particularly at the edges of objects, where significant visual features often reside. This noise can adversely affect the map's quality, especially in dynamic or complex environments like flight simulators.

To mitigate this limitation, researchers have developed methods that exploit the advantages of depth data from RGB-D cameras. For example, instead of solely relying on noisy point features, planar surface features inferred from depth information have been shown to provide a more robust basis for mapping. Planar features are generally more stable and less noisy than edges or corners, enabling better mapping accuracy. Gao and Zhang [124] proposed a method that selects and stabilizes depth values or planar points to improve SLAM system performance. Su et al. [125] demonstrated a technique for registering 3D point clouds using colour and depth data within the Kinect Fusion framework. Similarly, Li and Lee [126] developed a frame-to-keyframe edge-based RGB-D SLAM system that enhances pose estimation by combining foreground edge points with an intensity-assisted iterative closest point (IAICP) approach. Dai et al. [127] introduced Delaunay triangulation to analyze changes in edges formed by adjacent frames, improving feature point correlation and distinguishing between static and dynamic map elements.

In the context of augmented reality (AR) for flight simulators, RGB-D SLAM, leveraging depth data from devices such as the Microsoft HoloLens 2, provides a highly realistic and immersive simulation experience. By accurately modelling the 3D environment, RGB-D SLAM ensures precise spatial alignment between the pilot and the virtual cockpit, significantly enhancing the effectiveness of AR-based training.

Advantages

1. Improved Spatial Awareness: The integration of RGB and depth data from the HoloLens 2 cameras enables the creation of a more detailed and realistic 3D representation of the cockpit environment. This enhances users' depth perception and spatial awareness, crucial for practical AR-based training.

2. Real-Time Mapping: RGB-D SLAM's real-time ability to update environmental maps is particularly well-suited for dynamic simulation scenarios. In in-flight

simulators, where conditions and user interactions change continuously, this capability ensures a seamless and responsive user experience.

3. Enhanced Interaction: By providing an accurate 3D representation of the environment, RGB-D SLAM facilitates interactive and immersive experiences in flight simulators. For example, it supports intuitive interactions with virtual controls, displays, and realistic simulation of lighting and shadow effects in the virtual cockpit, further enhancing the trainee's immersion.

Limitations

1. High Computational Requirements: RGB-D SLAM is computationally intensive, requiring significant processing power to generate accurate maps in real time. This poses challenges for devices like the HoloLens 2, constrained by limited computational capacity.

2. Sensitivity to Lighting Conditions: RGB-D SLAM's accuracy heavily depends on the quality of depth data, which can be affected by varying lighting conditions. Changes in light intensity can introduce errors in the produced maps, undermining the system's reliability in specific scenarios.

3. Occlusions: RGB-D SLAM may struggle to capture accurate depth information in scenarios where objects obstruct or overlap in the camera's view. This limitation is particularly pronounced in cluttered or complex environments such as a cockpit, where occlusions can lead to incomplete or inaccurate maps.

3.2.3 ORB-SLAM (Oriented FAST and Rotated BRIEF SLAM)

ORB-SLAM is a state-of-the-art, monocular vision-based SLAM system that utilizes ORB (Oriented FAST and Rotated BRIEF) feature detection algorithms for real-time tracking and mapping. Initially developed by Rublee et al. [114], ORB feature detectors provide computational efficiency and robustness, making ORB-SLAM foundational for real-time SLAM applications. It has been widely adopted for tasks such as calculating camera trajectories and performing sparse 3D reconstructions using monocular, stereo, or RGB-D cameras, including applications in augmented reality (AR) for flight simulators.

According to Mur-Artal et al. [118], ORB-SLAM processes image sequences in real-time to estimate both camera pose and a sparse environment map. The system constructs the environmental map using keyframes and map points, with each keyframe storing a list of 2D features and their corresponding locations in the ORB-SLAM coordinate system. The algorithm's architecture includes three parallel threads: tracking, local mapping, and loop closure.

Tracking: This thread localizes the camera in real-time by matching features between incoming frames and the local map, minimizing re-projection errors.

Local Mapping: This thread updates the environmental map by managing keyframes and performing local Bundle Adjustment (BA) to optimize accuracy.

Loop Closure: This thread detects large loops in the trajectory to correct accumulated drift. Once a loop is identified, a global optimization step is performed using full BA to refine the map's structure and motion parameters.

To enhance ORB-SLAM's performance, Cai et al. [128] introduced a version incorporating affine transformations to improve feature extraction reliability. This enhancement addresses challenges like keyframe loss and limited feature detection under certain conditions. Despite its strengths, ORB-SLAM assumes a static environment, which can limit its applicability in dynamic scenarios like flight simulators. Recent advancements, including DynaSLAM [129] and other dynamic SLAM techniques [130], have mitigated these limitations by improving performance in environments with moving or changing objects.

Advantages

1. Real-Time Performance: ORB-SLAM excels in real-time mapping and localization, a critical requirement in flight simulators where timely decisions are essential.
2. Robustness: The system demonstrates resilience to lighting variations and can handle abrupt movements or changes in motion, making it suitable for dynamic environments.
3. Scalability: ORB-SLAM efficiently scales to handle large maps and complex environments, making it appropriate for large-scale simulations.
4. Open-Source Adaptability: As an open-source platform, ORB-SLAM can be customized and adapted to meet specific research and development needs, enhancing flexibility.

Limitations

1. Dependence on Keypoints: ORB-SLAM relies heavily on detecting and tracking keypoints. Its performance can degrade in scenarios with occlusions, poor lighting, or a lack of distinct features.
2. Limited Map-Building Capabilities: While effective at real-time tracking, ORB-SLAM's ability to create detailed maps is less advanced than SLAM systems designed for higher accuracy.
3. High Computational Requirements: The system requires significant computational resources to achieve real-time performance, posing challenges for resource-constrained devices like the Microsoft HoloLens.

3.2.4 LSD-SLAM (Large-Scale Direct Monocular SLAM)

LSD-SLAM is a direct monocular SLAM system that differs fundamentally from feature-based methods like ORB-SLAM by relying on pixel intensities rather than extracting distinct keypoints for tracking and mapping. This approach directly aligns images to track camera motion and generates semi-dense depth maps through pixel-wise stereo comparisons. These depth maps estimate the environment's geometry, enabling large-scale mapping and scale-drift correction, including loop closures [31]. Designed for efficiency, LSD-SLAM operates in real-time on CPUs and mobile devices,

making it a strong candidate for low-power systems such as the Microsoft HoloLens 2. Its capability to build large-scale maps through a pose graph of keyframes enhances its scalability compared to traditional monocular SLAM methods. LSD-SLAM's operation is structured around two primary stages:

Keyframe Selection: Keyframes are selected based on significant changes in visual information between successive frames, ensuring efficient map construction.

Direct Image Alignment: Instead of relying on feature matching, LSD-SLAM aligns images by directly comparing pixel intensities between the current frame and the keyframes, estimating the camera's motion through transformations in rotation and translation.

Several advancements have refined LSD-SLAM's capabilities to address its inherent limitations. Forster et al. [131] introduced SVO (Semi-Direct Visual Odometry), eliminating the need for loop closure detection and re-localization by focusing solely on motion estimation. This approach simplifies the computation process while maintaining accuracy. Similarly, Bergmann et al. [132] enhanced LSD-SLAM's robustness by integrating online photometric calibration, allowing the system to dynamically adjust to lighting variations, thereby improving direct visual odometry accuracy. Liu et al. [133] further refined this technique by incorporating sparse direct visual odometry, which tracks Shi-Tomasi corners in images to improve tracking and photometric calibration, making the system more robust in diverse conditions.

In the context of AR for flight simulators, LSD-SLAM's ability to operate in real-time and its compatibility with resource-constrained devices like the HoloLens 2 make it a viable choice. Its reliance on semi-dense maps ensures efficient mapping while providing the accuracy required for simulating realistic flight dynamics.

Advantages

1. Real-Time Performance: LSD-SLAM is specifically designed for real-time applications, making it highly suitable for flight simulators where immediate feedback and updates are critical.

2. High Accuracy: The direct approach to image alignment provides precise motion tracking, ensuring reliable performance in flight simulation dynamics.

3. Low Computational Requirements: LSD-SLAM is computationally efficient, requiring fewer resources than feature-based SLAM methods. This makes it particularly suitable for devices with limited processing power, such as AR headsets.

Limitations

1. Difficulty Handling Dynamic Environments: LSD-SLAM struggles with rapid movements or real-time environmental changes, limiting its accuracy in highly dynamic scenarios like flight simulators.

2. Sensitivity to Lighting Conditions: As LSD-SLAM relies heavily on visual data, variations in lighting or the presence of camera noise can significantly impact its accuracy and reliability.

3. Need for Accurate Initialization: The system requires a precise initial estimate of the camera's position and orientation, which can be challenging in flight simulator setups with complex and dynamic starting conditions.

4. Limited Robustness: In cases of insufficient or low-quality input data, LSD-SLAM is prone to tracking loss or inaccuracies in mapping, affecting its overall reliability.

3.2.5 VINS (Visual-Inertial SLAM)

Visual-inertial SLAM (VINS) is a hybrid technique that integrates visual data from cameras with inertial data from sensors such as gyroscopes and accelerometers. This fusion allows VINS to estimate the camera's pose and create a map of the environment with improved accuracy compared to vision-only SLAM methods. By combining these two data streams, VINS enhances motion tracking precision, enabling robust localization even in challenging conditions where visual input alone may be insufficient. As the camera navigates through the environment, the VINS algorithm dynamically updates the map, adding new features and removing outdated ones to maintain relevance.

A significant advantage of VINS over conventional monocular or stereo SLAM techniques is its ability to maintain accurate motion estimates in scenarios with limited or unreliable visual data, such as visually cluttered or obscured environments. This capability is particularly valuable in long-term mapping and tracking applications, where VINS can provide drift-free estimations of position and orientation over extended periods. Qin et al. [134] proposed a keyframe-based VINS framework that tightly integrates inertial measurements into the visual SLAM system, optimizing both re-projection errors and the cost function derived from the IMU.

Integrating visual and inertial information can be categorized into two approaches: loosely coupled and tightly coupled [134], [135]. In the loosely coupled approach, visual and inertial measurements are processed independently, with their outputs fused later. While this method is more straightforward, it may still suffer from drift, similar to monocular SLAM. Conversely, tightly coupled approaches incorporate inertial data directly into the state vector of the visual SLAM system, enabling simultaneous optimization of visual and inertial inputs. This integration corrects for drift and improves overall robustness and accuracy. For instance, Yin et al. [136] demonstrated a stereo visual-inertial SLAM system that combines stereo scene flow with IMU data for nonlinear optimization, significantly enhancing performance in dynamic environments.

In the context of augmented reality for flight simulators, the Microsoft HoloLens leverages a VINS-based SLAM algorithm that fuses high-quality visual data with inertial measurements from its onboard IMU. This enables the HoloLens to provide highly stable and accurate estimates of its position and orientation within the simulator environment, creating a more immersive and precise training experience. By integrating visual-inertial data, the HoloLens ensures seamless interactions between the user and the virtual cockpit, surpassing the capabilities of systems relying solely on visual or inertial inputs.

Advantages

1. High Precision: The fusion of visual and inertial data enables highly accurate estimations of location and orientation, enhancing tracking fidelity in complex and dynamic environments.
2. Robustness: VINS maintains accurate position estimates even in low-light conditions or during temporary losses of visual data, making it suitable for scenarios where visual input may be unreliable.
3. Real-Time Performance: VINS supports real-time processing, ensuring smooth and responsive navigation essential for virtual environments, including flight simulators, robotics, and AR applications.
4. Autonomy: VINS allows devices like the HoloLens to operate independently of external tracking systems, reducing reliance on additional infrastructure and enhancing usability.

Limitations

1. Sensitivity to Initial Conditions: The accuracy of VINS depends heavily on proper system calibration and initialization. Incorrect initial conditions can adversely impact tracking performance.
2. High Computational Requirements: The tight integration of visual and inertial data demands significant computational resources, which can be challenging for low-power devices such as mobile AR systems and wearables like the HoloLens.
3. Residual Drift Over Time: While VINS reduces drift compared to visual-only SLAM, accumulated errors over extended periods can still result in deviations in position and orientation estimations.

3.2.6 Deep Learning SLAM Methods

Deep learning has demonstrated significant advancements in image processing tasks, particularly in data fitting, feature extraction, and spatial transformations. These capabilities have driven remarkable progress in applications like image recognition [137], segmentation [138], object detection [139], and classification [140], where deep learning has consistently outperformed traditional algorithms.

In the context of SLAM, integrating deep learning addresses several limitations of traditional visual odometry and scene recognition approaches, which often rely on manually designed features. By leveraging data-driven learning, deep learning enhances an agent's perceptual and cognitive capabilities, enabling high-level semantic understanding and the development of a knowledge base tailored to specific environments. This evolution from geometry-based SLAM methods to those enhanced by deep learning has expanded the possibilities of visual SLAM, mainly through advancements in supervised and unsupervised deep learning techniques. These methods have been applied to critical SLAM tasks such as visual odometry [141] and loop closure [142], which identify and correct trajectory estimation errors, significantly improving system accuracy and adaptability. For instance, Geng et al. [143] proposed an unsupervised end-to-end learning framework that integrates ORB-SLAM to improve camera motion

and monocular depth estimation. Similarly, Li et al. [144] utilized SegNet, a semantic segmentation network, to differentiate dynamic objects in an environment. Zhang et al. [145] employed YOLO [146] for object detection and used a semantic map to filter out dynamic feature points, enhancing the overall system's precision.

Further innovations include Tateno et al. [147], who developed a dense monocular reconstruction method by combining CNN-predicted dense depth maps with direct monocular SLAM, and Li et al. [148], who introduced UnDeepVO, an unsupervised framework for training stereo images to estimate camera pose and create maps from monocular data. Another significant contribution is SfM-Net, developed by Vijayanarasimhan et al. [149], which uses geometrically generated images for training to extract 3D structures, ego-motion, and segmentation from video data.

Deep learning-based SLAM techniques hold significant promise in the context of augmented reality (AR) for flight simulators. Dynamic environments, such as cockpits in flight simulators, require high accuracy, robustness, and adaptability. By incorporating these techniques into AR devices like the Microsoft HoloLens, systems can achieve enhanced navigation, object recognition, and scene understanding. These improvements enable more realistic simulations, fostering immersive and effective training experiences.

Advantages

1. Fusion of SLAM with Deep Learning: Deep learning-based SLAM seamlessly integrates with traditional SLAM algorithms. Neural networks can refine SLAM outputs by improving the detection of critical features or reducing noise, resulting in enhanced accuracy.

2. Improved Accuracy: Deep learning enhances the accuracy of maps generated by SLAM systems. Estimating object depth through data-driven training enables creating more precise and reliable 3D maps.

3. Increased Robustness: Deep learning models improve SLAM robustness, particularly in difficult situations like low-light settings or rapid movements. They can be trained to detect and correct errors in SLAM outputs, enhancing system reliability in dynamic scenarios.

Limitations

1. Data Collection and Annotation: Deep learning algorithms require large amounts of labelled data for training. Collecting and annotating such data in flight simulator environments can be time-consuming and labour-intensive.

2. High Computational Requirements: Deep learning models are computationally intensive, posing challenges for devices like the HoloLens, which have limited processing power compared to high-performance systems.

3. Model Size: Deep learning models are often large, requiring significant storage capacity. Due to hardware constraints, deploying such models on devices like the HoloLens is challenging.

4. Training Costs: Training deep learning algorithms can be resource-intensive and expensive, potentially limiting their feasibility for commercial flight simulator applications with significant budget constraints

3.2.7 Experimental Setup

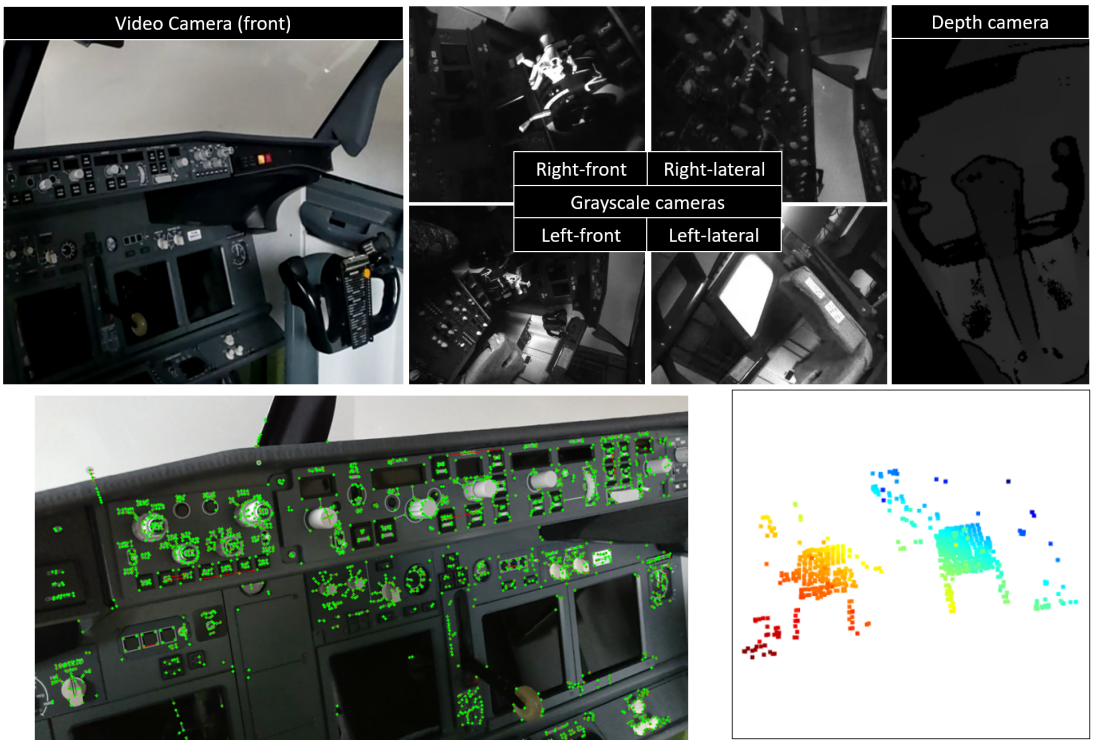
The experimental setup was designed to evaluate the performance of simultaneous localization and mapping (SLAM) techniques within a simulated flight cockpit environment using the HoloLens 2. The primary objective was to assess the ability of the SLAM methods to perform real-time mapping and pose estimation, with a particular focus on feature extraction and localization accuracy in a controlled cockpit setting. The HoloLens 2 was selected for its advanced sensor suite, which includes depth sensors, an Inertial Measurement Unit (IMU), and grayscale cameras. These sensors were utilized in Research Mode (RM) to capture real-time data streams during the experiment. A Python interface facilitated communication with the HoloLens 2 server, enabling command transmission, device configuration, and continuous sensor data streaming, including depth, IMU measurements, and grayscale images.

The experimental process began with real-time data acquisition from the cockpit environment. Video streams and sensor data from the HoloLens 2 cameras were collected by connecting to the device server using the open() method. Continuous frames were captured during this phase through repeated calls to the next-get-packet() method, ensuring a steady and consistent data flow. Once sufficient data was collected for analysis, the connection was closed using the close() method [150]. This step ensured the integrity of the dataset and streamlined subsequent processing.

After data acquisition, one of the SLAM techniques under evaluation was applied to the captured video stream to extract key features from the cockpit environment. This process involved analyzing the video data for patterns, structures, and landmarks critical to constructing an accurate map of the environment. The selected SLAM method was also responsible for performing precise pose estimation, leveraging the extracted features to localize the HoloLens 2 within the cockpit. The choice of the specific SLAM method was based on its demonstrated performance in the simulated environment, as detailed in the results section. This setup allowed for a comprehensive assessment of SLAM performance in a flight simulation context, ensuring the experimental design was rigorous and replicable.

3.3 Results and Discussion

In this study, the HoloLens 2 was employed to execute simultaneous localization and Mapping (SLAM) algorithms within a flight simulator environment. The system provided real-time live-streaming video from multiple cockpit cameras. Utilizing the HoloLens 2's Research Mode (RM), as shown in Figure 1, data was captured from depth sensors, inertial measurement units (IMU), and grayscale cameras. The connection between the Python library and the HoloLens 2 server facilitated the transmission of commands, data reception, and subsequent analysis using complementary libraries.



(A) Feature detection from the cockpit.

(B) 3D map of the cockpit.

FIGURE 3.1: Top image shows the simultaneous data capture from all HoloLens 2 sensors. The left image (a) displays feature detection, while the right image (b) depicts the 3D map of the cockpit.

The primary objective of integrating SLAM techniques within the flight simulator cockpit was to evaluate their adaptability in detecting patterns and structures, constructing an accurate map of the environment, and enabling reliable pose estimation. The results of these experiments, summarized in Table 1, highlight the comparative performance of various SLAM techniques under the controlled conditions of the flight simulator environment.

Our experiments revealed significant variations in the performance of the SLAM techniques evaluated. The RGB-D SLAM method performed moderately well due to the flight simulator's structured nature, simplifying feature detection. However, its limited robustness in handling dynamic environmental changes constrained its performance.

Among the visual-based techniques, ORB-SLAM and LSD-SLAM exhibited contrasting behaviours. ORB-SLAM struggled in this environment due to its sensitivity to lighting variations and occlusions, which made object distinction challenging. In contrast, LSD-SLAM achieved high mapping quality in smaller environments but demonstrated poor accuracy in the large-scale setting of the flight simulator. This limitation, combined with its reduced efficiency in dynamic environments, resulted in overall suboptimal performance for LSD-SLAM.

VINS, which integrates data from both visual and inertial sensors, demonstrated high accuracy and robustness. This hybrid approach allowed VINS to adapt effectively

Methods	Map Quality	Localization Accuracy	Computational Cost	Robustness	Efficiency	In Flight Simulator
RGB-D	Moderate	Good	High	Poor	Fair	Moderate
ORB	Low	Good	Moderate	Good	Good	Poor
LSD	High	Excellent	Low	Fair	Fair	Poor
VINS	High	Excellent	High	Excellent	Good	Good
DL	Good	Excellent	Very High	Good	Low	Good

TABLE 3.1: Comparison of SLAM methods in the flight simulator environment

to dynamic environments, performing well without additional sensors. Its ability to maintain real-time performance made it particularly suitable for flight simulators, where rapid adaptability is critical. Deep learning (DL)—based SLAM techniques also showed promising results. These methods excelled in learning and adapting to varying environments by extracting meaningful features from sensor data. However, the computational cost associated with DL approaches was significantly higher than that of other techniques, posing challenges for their implementation on devices like the HoloLens 2, which are constrained by limited processing power.

Our findings indicate that VINS and deep learning-based SLAM methods are the most reliable for achieving high accuracy and robustness in the flight simulator environment. While RGB-D SLAM demonstrated moderate performance in controlled conditions, it was less effective in dynamic scenarios. Similarly, ORB-SLAM and LSD-SLAM struggled to handle the simulator's large-scale and changing nature, limiting their applicability in this context.

3.4 Conclusion

This chapter presented an in-depth evaluation of Simultaneous Localization and Mapping (SLAM) methods within an augmented reality (AR) flight simulator environment, leveraging the advanced sensor capabilities of the HoloLens 2. The integration of SLAM algorithms facilitated the extraction of structural features from the cockpit and enabled real-time estimation of position and orientation. These capabilities provided precise movement tracking and the construction of accurate, dynamic maps essential for immersive and responsive AR applications. The comparative analysis of SLAM techniques revealed critical trade-offs between computational cost, accuracy, robustness, and adaptability in dynamic environments such as flight simulators. Hybrid approaches, such as Visual-Inertial SLAM (VINS), demonstrated superior robustness and adaptability by fusing visual and inertial sensor data. Deep learning-enhanced SLAM methods also highlighted significant potential in addressing challenges like feature detection and dynamic object tracking. However, these approaches remain constrained by computational demands and resource limitations.

Future studies will focus on enhancing the efficacy of Visual Inertial SLAM and upgrading it based on the findings of this research. Using augmented reality goggles,

we will demonstrate how the proposed techniques can create more precise and dependable algorithms for feature detection, tracking, and aligning virtual objects in the real world. This will help to reduce the incidence of tracking failures. We will present the development of better sensor fusion techniques by integrating with other sensors and integrating information from various sensors so the system can better understand its environment and its position. This model can aid in improving the SLAM system's accuracy and help it better forecast aircraft behaviour under various flight conditions. In augmented reality, one of the significant challenges is the ability to operate in a dynamic and unstructured environment. We will create an algorithm capable of handling challenging conditions in various environments, such as low light, moving objects, and dynamic scenes. Additionally, we will apply SLAM methods to reduce the latency caused by occlusions, where real-world objects obscure virtual objects. SLAM techniques can help to ensure that virtual items are perfectly aligned with the actual world, even when the user moves and the environment changes, by continually updating the device's position and orientation.

The insights gained from this study directly contribute to the broader objectives of this thesis. This chapter lays a critical foundation for developing more robust and scalable augmented reality systems by identifying the strengths and limitations of SLAM algorithms. These findings inform the subsequent chapters by highlighting the technical and computational challenges inherent in AR-based tracking and mapping systems, guiding the exploration of novel optimization techniques and integrating hybrid methods. Moreover, the conclusions drawn here emphasize the transformative potential of SLAM in AR applications, particularly in training and simulation contexts. Achieving accurate, real-time alignment of virtual objects with real-world environments is vital for creating compelling and immersive training systems. This chapter's outcomes are a stepping stone for further advancements, including optimizing SLAM performance for resource-constrained AR devices and improving scalability for complex, dynamic environments.

This chapter advances the state-of-the-art understanding of SLAM in AR flight simulator environments and provides a practical framework for implementing SLAM-based AR systems across various domains. By addressing the challenges of environmental complexity, computational efficiency, and sensor variability, this work strengthens the thesis' aim to develop innovative AR technologies for real-world applications, including aviation training and beyond.

4 Analysis of Marker and SLAM-based Tracking for Advanced Augmented Reality (AR)-based Flight Simulation

4.1 Introduction

Integrating augmented reality (AR) into flight simulation represents a transformative advancement in pilot training methodologies. It provides engaging and interacting settings that bridge the gap between traditional training tools and real-world complexity. Unlike conventional simulation methods, AR-based flight simulators seamlessly blend digital overlays with the physical world, enabling pilots to engage with complex scenarios in risk-free conditions. This innovation significantly reduces training costs, fuel consumption, and safety concerns while enhancing learning outcomes. Conventional flight simulators, such as Full-Flight Simulators (FFS) and Computer-Based Training (CBT), have historically formed the backbone of pilot training. These tools provide controlled environments to develop essential skills, yet they often fail to replicate actual flight conditions' dynamic, immersive qualities. This limitation hinders the effective transfer of skills from simulation to reality, as noted in prior studies [151]. Bridging this gap is critical for advancing pilot proficiency, situational awareness, and decision-making capabilities.

Augmented reality addresses these challenges by superimposing digital information onto real-world environments. This integration enhances training realism, enabling pilots to simulate complex flight scenarios without physical risks [152]. Two primary tracking methods—marker-based tracking and Simultaneous Localization and Mapping (SLAM)—are the foundation of AR systems. While marker-based tracking offers high precision in controlled settings, SLAM provides adaptability in dynamic environments, making it essential in diverse AR training applications. Understanding their comparative effectiveness is vital for developing tailored training solutions.

4.1.1 Marker-Based Tracking in AR Flight Simulator

Marker-based tracking is a widely utilized method in AR systems, where distinct physical markers are employed to determine the camera's or AR device's position and orientation. These markers serve as reference points, enabling precise alignment of

[0]Publication note: This chapter is based on the peer-reviewed article [37]. https://doi.org/10.1007/978-3-031-63783-4_16.

virtual elements within the real-world environment. Marker-based tracking systems often rely on visual identifiers, such as fiducial markers [153], QR codes [154], and custom images, to provide reliable and accurate positional data, as illustrated in 4.1. This method is particularly beneficial in controlled environments where high precision is required, such as in cockpit simulations for pilot training.

Marker-based tracking, with its high accuracy and ease of implementation, offers significant benefits in flight training. Its capability to maintain stable and precise overlays makes it ideal for tasks that require exact alignment of virtual objects, such as procedural training and instrument handling in flight simulators. Various implementations of marker-based tracking have demonstrated its effectiveness in enhancing training experiences by overlaying virtual controls and displays onto physical cockpits, thus preserving the tactile feedback essential for pilot training.

Research on marker-based tracking has explored different types of visual markers to optimize AR applications. For instance, Naimark and Foxlin's work on concentric rings demonstrated the use of geometric patterns to improve tracking precision, which has benefited applications requiring detailed spatial accuracy [155]. Similarly, Deineko et al. highlighted the efficiency of QR codes, known for their quick readability and data capacity, which can enhance AR tracking in scenarios where rapid data exchange is needed [154]. Platforms such as Vuforia have enabled the customization of markers, allowing the use of specific images as reference points for various applications. This adaptability has been particularly useful in flight training, where personalized markers can match specific cockpit instruments, providing a seamless and realistic training experience. Ribeiro et al. demonstrated printed markers for UAV pilot training, creating an accessible and cost-effective solution by overlaying virtual flight data on a physical interface [156].

Despite its advantages, marker-based tracking has limitations. The requirement for physical markers can be restrictive, particularly in dynamic environments where the markers might be occluded or challenging to set up. Wallace et al. utilized ArUco tags to overlay virtual gauges onto physical instrument panels, enhancing immersion while addressing some setup challenges. However, occlusion and marker placement issues remained, especially in scenarios involving complex movements [157], [158]. Building on these findings, this study explores novel marker configurations and optimization techniques tailored for cockpit environments, aiming to reduce occlusion and improve setup flexibility.

The insights gained from marker-based tracking serve as a foundation for evaluating and comparing it with SLAM-based approaches in subsequent sections, ultimately contributing to developing a robust AR tracking system for immersive flight training.

4.1.2 SLAM-Based Tracking in Augmented Reality

Simultaneous Localization and Mapping (SLAM) represents a more flexible approach to AR tracking, particularly suited for environments where physical markers are impractical [159]. Unlike marker-based tracking, SLAM systems dynamically map the environment while simultaneously tracking the position of the AR device within that environment. This adaptability allows for real-time adaptation to environmental changes, making SLAM particularly advantageous in scenarios that require training under unpredictable or dynamic conditions.

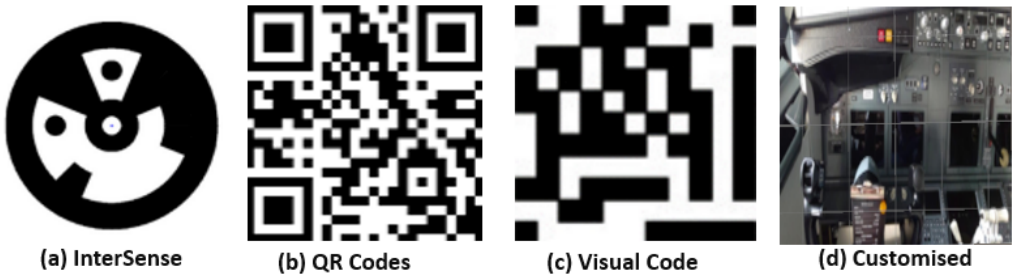


FIGURE 4.1: Various AR Tracking Markers

To create detailed environmental maps, SLAM technology integrates various sensors, including cameras, LiDAR, and Inertial Measurement Units (IMUs). Devices like the Microsoft HoloLens 2 utilize advanced SLAM algorithms to overlay virtual elements seamlessly, enabling users to navigate and interact with complex AR environments. The adaptability of SLAM makes it ideal for training applications that simulate real-world variabilities, such as emergency procedures or complex flight maneuvers.

The need for adaptive and robust tracking solutions has driven the development of SLAM-based tracking. Early approaches, such as MonoSLAM, laid the groundwork for visual SLAM by using monocular cameras to build real-time maps, offering the first steps towards AR applications without reliance on physical markers [28]. Subsequent innovations, including ORB-SLAM and LSD-SLAM, have improved upon these foundations by incorporating advanced feature detection algorithms that enhance tracking accuracy and environmental mapping [31], [118]. These systems employ feature extraction algorithms such as ORB or SURF to identify key points, while sensor fusion integrates data from IMUs and cameras to improve accuracy during motion tracking.

Li et al. proposed a novel SLAM system, DP-SLAM, to address the challenges of dynamic environments where most SLAM systems fail due to their static-world assumption. Their system leverages a moving probability propagation model to detect dynamic keypoints by estimating the likelihood of a keypoint being located on moving objects. By combining geometry constraints and semantic segmentation within a Bayesian probability estimation framework, DP-SLAM effectively tracks dynamic keypoints. Integrated into the front-end of ORB-SLAM2, their approach acts as a preprocessing stage to filter out keypoints associated with moving objects. [160]. Their system includes a background inpainting process to handle occluded areas, which enhances applications in virtual and augmented reality. Experimental results on the TUM RGB-D dataset and custom sequences demonstrate that DP-SLAM significantly improves the performance of state-of-the-art SLAM systems in various challenging scenarios. Safi and Chung's system utilizes real-time data acquisition to identify cockpit elements, improving the accuracy of learning tools by reducing user errors and enhancing overall efficiency [161]. Although these advancements have demonstrated the potential of SLAM in dynamic environments, challenges such as computational demands and accuracy constraints in confined spaces remain a critical focus for ongoing research.

Integrating additional sensors, such as IMUs, has been explored to provide more stable and reliable tracking. Recent research has also looked into integrating deep learning with SLAM to improve the system's ability to function in environments with minimal visual features or high levels of dynamic change [162]. For instance, combining machine learning with SLAM systems enables adaptive feature extraction and error correction, further enhancing their performance in dynamic AR scenarios.

Furthermore, integrating SLAM in devices like the HoloLens 2 has shown promise for flight training by enabling pilots to interact naturally with their training environment without the constraints of physical markers. This has allowed for more comprehensive simulations, encompassing various scenarios, including emergency procedures and complex navigational tasks. However, as Nwobodo et al. noted, challenges are still related to computational demands and accuracy in confined spaces, which remain a critical focus for ongoing research [16].

Building on these advancements, this study evaluates the performance of SLAM-based tracking systems in flight simulator environments, focusing on their adaptability and accuracy under constrained conditions. The insights gained will inform the development of hybrid AR training systems, leveraging the strengths of both marker-based and SLAM-based approaches to create an optimized tracking framework for pilot training. As SLAM technology evolves, its integration with machine learning and advanced sensors promises to overcome current limitations, paving the way for more immersive and effective AR training systems.

This chapter introduces the foundational concepts of AR-based flight simulation and explores the critical role of tracking technologies. Marker-based tracking and SLAM are analyzed, focusing on their uses, benefits, and drawbacks within the context of pilot training. By incorporating insights from existing research and experimental findings, this chapter contributes to the broader objectives of the thesis by addressing critical gaps in AR-enabled flight simulation and paving the way for more immersive, effective, and scalable training methodologies.

4.2 Methods

This chapter outlines the comprehensive methodology employed to evaluate the effectiveness of marker-based and SLAM-based tracking technologies in augmented reality (AR) flight simulation training using the Microsoft HoloLens 2. The methodological framework includes a systematic approach to camera calibration, marker design, and the establishment of mathematical relationships between the detection range and the intrinsic parameters of the HoloLens 2 camera. These elements are integral to ensuring the precision and reliability of AR tracking in flight simulation, enabling a robust and immersive training environment.

The research framework is designed to assess two primary tracking methods—marker-based and SLAM-based—through controlled experiments in a simulated cockpit environment. This study compares the two techniques by analyzing performance metrics such as localization accuracy, map quality, computational cost, and robustness under dynamic conditions. Each methodological step is tailored to address the unique challenges of AR tracking in flight simulation.

Camera Calibration

Accurate superimposition of virtual elements onto the real world is critical for the efficacy of AR simulation training, particularly in scenarios as demanding as flight simulation. This precision is fundamentally dependent on robust camera calibration. As illustrated in Figure 4.2, camera calibration facilitates the accurate transformation of 3D world coordinates $\mathbf{E}_w$ into 2D image points $\mathbf{V}_c$, thereby ensuring the seamless alignment of virtual content with the physical environment.

Camera calibration plays a pivotal role in ensuring the realism of AR-based flight simulations, as it directly affects the alignment of virtual overlays with the physical environment. Precise calibration is critical for applications such as pilot training, where even minor misalignments can significantly impact the effectiveness of the simulation. This process involves deriving intrinsic and extrinsic parameters using standardized methods, such as checkerboard calibration, which uses multiple views of a known pattern to solve for camera characteristics. Tools like OpenCV or proprietary SDKs, such as the HoloLens 2 calibration toolkit, are often employed for this purpose.

The calibration process involves determining both the intrinsic parameters—focal lengths f_x, f_y and the principal point coordinates x_0, y_0—and the extrinsic parameters, which describe the camera's pose $[R|t]$ relative to the environment. Understanding these parameters is essential for AR systems to consistently track markers and reference points, thereby ensuring precise integration of digital overlays. Key parameters such as the marker size S_M, distance D from the camera, and the field of view (FOV) are fundamental in defining the tracking range and optimizing the detection thresholds for seamless AR integration, especially in the constrained spaces of cockpit training modules. Incorporating these calibration parameters ensures the realistic rendering of virtual content, crucial for maintaining the immersive nature of pilot training modules, thus enhancing the learning and skill acquisition process. The geometric principles underpinning these calibrations are depicted through the convergence of lines at the principal point in Figure 4.2, emphasizing the importance of precise configuration for optimal AR performance.

A critical aspect of the calibration process is estimating the camera's optimal detection range. The effectiveness of marker-based tracking, particularly in AR devices such as the HoloLens 2, relies on carefully calibrated detection range settings. This range is instrumental for applications such as flight simulation, where the size and placement of printed markers must be optimized for reliable recognition by the onboard camera. The fundamental relationship governing this detection range can be expressed as follows [163]:

$$D = \frac{S_M}{s_M}, \tag{4.1}$$

where S_M is the printed marker size, and s_M is the normal image plane marker size at a distance $D = 1$, as illustrated in Figure 4.2.

This equation establishes the fundamental principle for determining optimal marker placement relative to the HoloLens 2 camera, ensuring that markers are consistently within the device's field of view and can be reliably detected. Extending this,

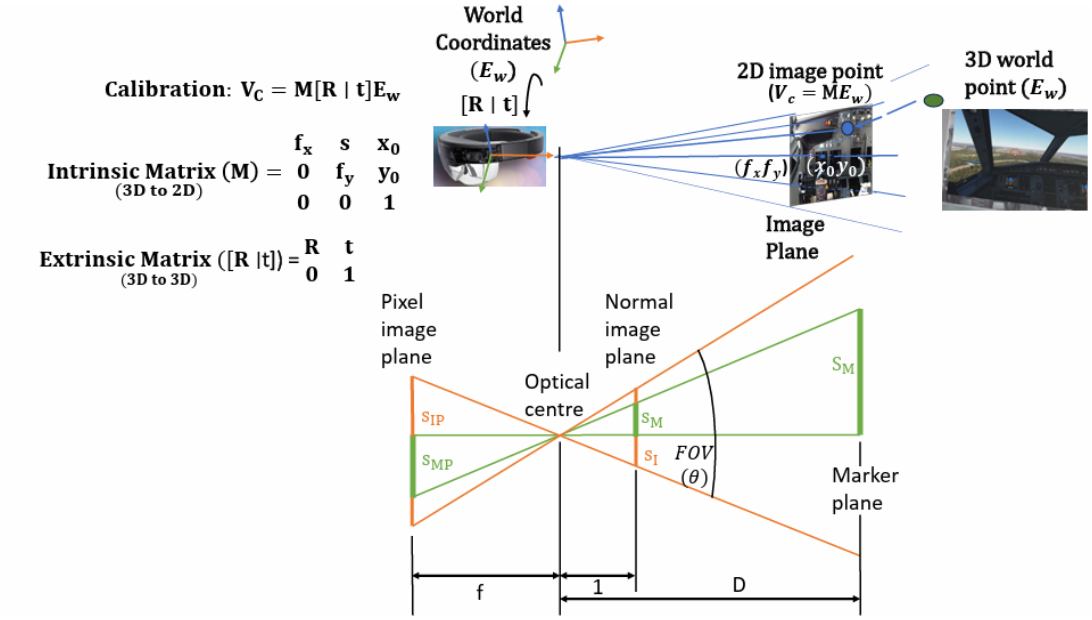


FIGURE 4.2: The figure depicts the AR camera calibration, showing 3D world coordinates $\mathbf{E}_w$ projection onto 2D image points $\mathbf{V}_c$, including the camera's intrinsic matrix M and extrinsic parameters $[R|t]$. It also shows the principal point, FOV lines, marker size S_M, and distance D from the camera. (The lowest part of the figure is adapted from [163])

the detection range D can be further refined as:

$$D = \frac{S_M \cdot s_I}{s_{IP} \cdot s_{MP}} = \frac{S_M \cdot f}{s_{MP}} \tag{4.2}$$

where

$$s_I = 2 \cdot \tan\left(\frac{\theta}{2}\right), \quad s_{IP} = 2f \cdot \tan\left(\frac{\theta}{2}\right).$$

Here, f denotes the camera's focal length, and θ the field of view angle. From this, it can be inferred that for given parameters of the HoloLens 2 (f and s_{MP}), the detection range is directly proportional to the printed marker size S_M. The parameter s_{MP} can be specifically obtained for different configurations of the HoloLens 2.

While these equations provide a robust theoretical framework, practical implementation can face challenges such as lens distortion, ambient lighting variability, and dynamic marker movement. These factors must be carefully accounted for for optimal calibration and marker tracking reliability. This highlights the importance of iterative testing and validation of the calibration process in environments that mimic real-world conditions.

Equations (4.1) and (4.2) are vital for determining the detection range D and ensuring accurate marker tracking within the simulation environment. The practical application of these theoretical principles is demonstrated through the marker-tracking process, as shown in Figure 4.3.

Marker Tracking Workflow

The marker tracking workflow is a critical component of augmented reality (AR) systems, particularly in applications like flight simulation, where precise alignment of virtual and real-world elements is paramount. This workflow begins with capturing frames containing markers, which are processed through a sequence of operations: detection, recognition, filtering, and pose estimation.

The detection phase identifies potential marker regions within the frame, often leveraging image processing techniques [164] or deep learning algorithms [165] to improve robustness against variations in lighting or perspective. In the recognition phase, the system decodes the unique patterns or identifiers within the markers, ensuring correct mapping to their corresponding virtual objects [166]. Filtering mechanisms are then applied to reduce noise and eliminate false positives, enhancing the accuracy of the subsequent steps [167]. Finally, pose estimation calculates the spatial orientation and position of the markers relative to the camera using geometric principles and algorithms such as the Perspective-n-Point (PnP) method [168].

This sequence of operations enables the accurate overlay of virtual objects onto the captured frames, ensuring proper alignment with the physical environment. Such precision is vital for maintaining the fidelity of AR elements and delivering an immersive training experience. For instance, in flight simulation, where pilots rely on visual and spatial cues, any misalignment of virtual overlays can compromise the training's effectiveness.

Despite its robustness, marker tracking faces challenges such as marker occlusions, motion blur, or environmental variations (e.g., glare or shadows) [169]. These issues necessitate adaptive algorithms and iterative calibration to ensure reliable performance in dynamic conditions. Additionally, optimizing marker design, such as adjusting their size and contrast, can enhance detection under constrained cockpit environments [170].

The marker tracking workflow described here underscores its significance in achieving high-fidelity AR experiences, contributing to the overarching goal of realistic and effective pilot training in simulation environments.

SLAM-Based Tracking Process

The application of SLAM (Simultaneous Localization and Mapping) technology in the HoloLens 2 is illustrated in Figure 4.4. SLAM-based tracking is initiated through initialization, feature processing, and continuous tracking phases, complemented by mapping and loop detection functionalities [13]. These processes collectively enable the AR system to dynamically map the environment while tracking the position of the HoloLens 2 within it.

The SLAM process begins by capturing environmental features, such as keypoints and edges, using RGB-D sensors and inertial measurement units (IMUs) [85]. These features are processed to establish a baseline map, utilizing algorithms like ORB-SLAM2 [171] or visual-inertial SLAM [172] for real-time tracking. As the user moves, the system continuously updates the map and maintains spatial awareness, with loop detection ensuring that previously mapped regions are accurately recognized. This loop closure mechanism minimizes drift and enhances the stability of the generated

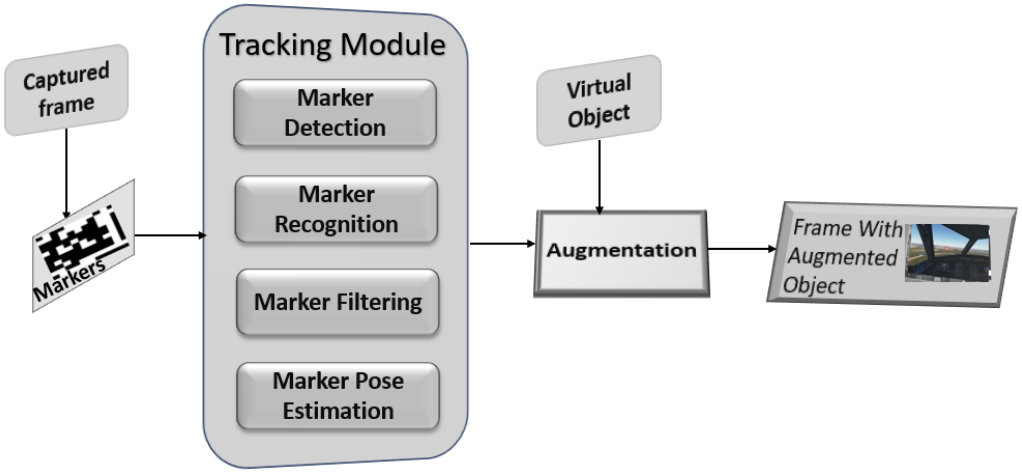


FIGURE 4.3: Flowchart depicting the marker tracking process, from initial frame capture to augmentation.

map, which is particularly critical for applications requiring high positional accuracy [173].

The SLAM system's mapping and tracking capabilities are instrumental for AR-based flight simulation training. Pilots navigate complex virtual scenarios superimposed on real-world environments, and consistent alignment of virtual elements is imperative for practical and immersive training experiences. The integration of SLAM allows the AR system to adapt dynamically to unstructured environments, ensuring seamless user interaction even in scenarios with limited predefined markers [174].

The practical integration of these methodologies was tested within an experimental framework designed to assess the robustness and applicability of marker-based and SLAM-based tracking for AR flight simulation [175].

These findings underscore the complementary nature of SLAM-based and marker-based tracking in AR applications, suggesting a hybrid approach for future flight simulation systems. This research advances the state of the art in AR tracking and offers a foundation for further exploration of adaptive tracking methodologies in high-stakes training environments.

4.2.1 Experimental Setup

This study investigates the transformative potential of augmented reality (AR) technologies in enhancing flight simulation training, specifically focusing on improving the execution of landing manoeuvres and comprehension of cockpit instrumentation. The research builds on the theoretical foundation of AR-based training, addressing limitations in traditional marker-based tracking while leveraging the adaptive strengths of SLAM technologies. By integrating these methodologies, the study aims to advance contemporary pilot training techniques through an innovative dual-tracking approach.

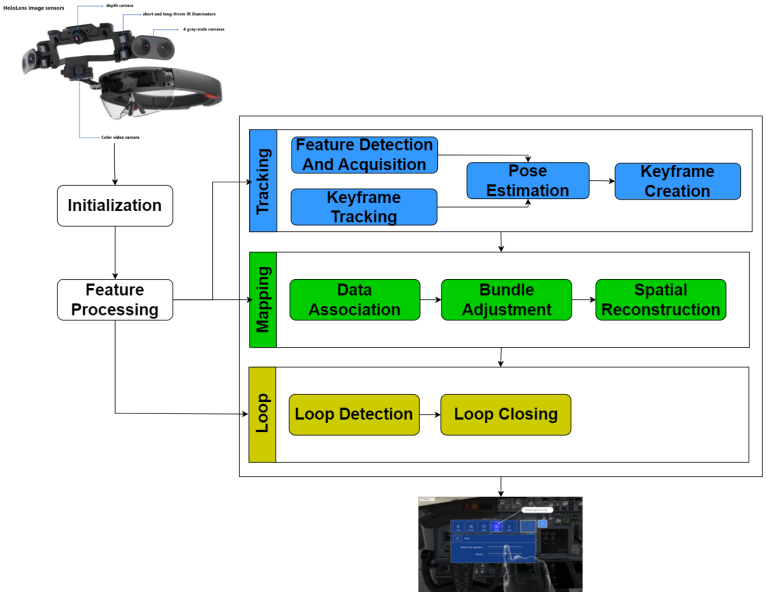


FIGURE 4.4: Flowchart illustrating the SLAM process in HoloLens 2, including initialization, feature processing, tracking, mapping, and loop detection phases.

4.2.2 Experimental Framework

The experimental framework was developed using the Microsoft HoloLens 2 and Unity software to create a state-of-the-art AR simulation environment. This system blends real-world cockpit settings with interactive virtual flight instruments, offering a high-fidelity training experience. The integration was achieved through Unity's AR Foundation and the Mixed Reality Toolkit (MRTK), enabling seamless interactions between physical and digital elements. This dual-tracking approach ensures a robust and immersive training environment.

The experimental design included 10 participants with no flight experience but varying levels of AR experience. This selection evaluated the system's usability and performance across various user backgrounds, ensuring comprehensive system testing. Each participant wore the HMD and performed simulated landing manoeuvres and cockpit instrumentation tasks.

The experiment incorporated a combination of marker-based and SLAM tracking technologies to address the specific needs of pilot training. These complementary methods were designed to leverage the precision of marker-based tracking for static elements and the adaptability of SLAM tracking for dynamic environments.

4.2.3 Marker-Based Tracking

The marker-based system strategically utilized high-contrast QR codes and custom-designed markers within the cockpit environment. These markers were printed in three distinct sizes, optimized for different training scenarios:

- S_{M1}: 10 cm $\times$ 10 cm for close-range instruments.

- S_{M2}: 16.2 cm $\times$ 16.2 cm for mid-range elements.
- S_{M3}: 29 cm $\times$ 20.4 cm for broader spatial calibration.

Equation (3.2) guided the positioning of these markers, ensuring they remained within the HoloLens 2 camera's effective detection range. Calibration included adjustments for interpupillary distance (IPD) and sensor optimization to enhance tracking accuracy. While marker-based tracking provided reliable precision, its limitations in dynamic conditions necessitated using SLAM-based tracking.

4.2.4 SLAM-Based Tracking

The SLAM-based tracking system was implemented to map the cockpit environment dynamically. This approach allowed the system to create a real-time three-dimensional map of the cockpit, enabling precise projection of virtual flight instruments and instructional overlays. The MRTK configurations were fine-tuned to optimize environmental understanding and computational efficiency, ensuring seamless performance during dynamic head movements. SLAM's adaptability addressed challenges like marker occlusion and varying lighting conditions, providing a reliable and continuous tracking experience.

4.2.5 Evaluation Metrics

To evaluate the performance of the tracking systems, several key performance indicators (KPIs) were defined:

- **Overlay Accuracy:** The alignment precision of virtual elements with the physical cockpit environment, measured in millimetres.
- **Overlay Precision:** The repeatability and consistency of virtual element alignment across multiple trials, quantified by standard deviation.
- **System Responsiveness:** The speed and stability of the tracking systems, quantified as latency (milliseconds) and frame rate (frames per second).
- **User Experience:** Participants' perception of immersion, realism, and ease of interaction during the training sessions, captured through post-experiment surveys using a 5-point Likert scale.

These metrics were chosen to evaluate the system's ability to provide precise, stable, and responsive AR overlays, addressing the study's objective of enhancing pilot training. Comparative analysis will be conducted to identify the strengths and weaknesses of each approach.

The primary aim was to enhance the efficiency of pilot training by integrating AR technologies within flight simulation, representing a significant advancement in contemporary training methodologies. By enabling pilots to engage with highly interactive and realistic scenarios, AR is poised to improve learning outcomes, offering enriched and dynamic training experiences beyond traditional simulation tools' limitations. Through this experimental approach, the study aims to contribute to developing more effective, engaging, and adaptive training systems in aviation. Figure 4.5 depicts the experiment setup for marker and SLAM tracking in unity.

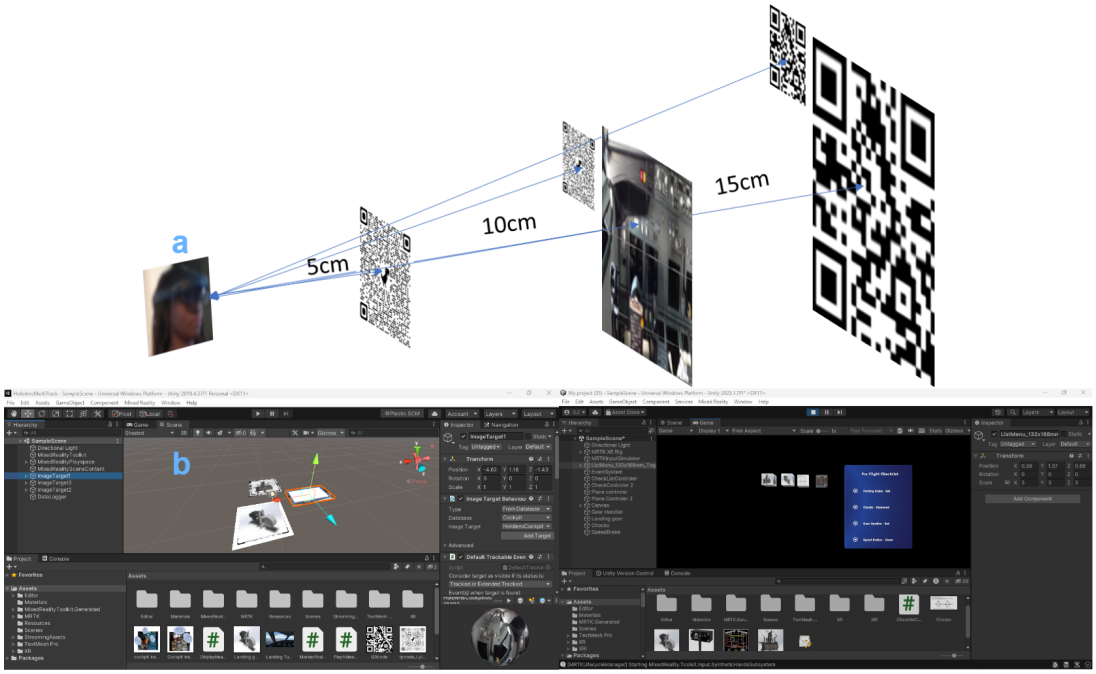


FIGURE 4.5: (a) Markers at varying distances for HoloLens 2 tracking calibration. (b) Unity interface setting up a simulated cockpit with marker detection and SLAM configuration for AR development. These visual elements illustrate the dual-tracking approach used in this study.

4.3 Result

This chapter presents the outcomes of a comparative analysis between marker-based tracking and Simultaneous Localization and Mapping (SLAM) technologies on the Microsoft HoloLens 2, specifically within the context of advanced flight simulation training. The study evaluated each technology under various detection ranges, focusing on precision, reliability, and adaptability.

The marker-based tracking system demonstrated high positional accuracy within a range of 5 cm to 15 cm from the HoloLens 2 camera, especially under well-lit conditions. This accuracy was particularly effective for tasks that require precise overlay alignment, which is essential for procedural training applications.

SLAM technology exhibited adaptability, leveraging GPS and gyroscopes to map and project virtual information onto diverse surfaces without needing pre-placed markers. This flexibility enhanced spatial awareness during training scenarios, although it showed slightly lower positional accuracy than marker-based systems.

The study further investigated how the size of printed markers affects the detection range. Larger markers extended the effective detection range, which is crucial for maintaining precision over greater distances. Figures 4.6 and 4.7 illustrate this dependency.

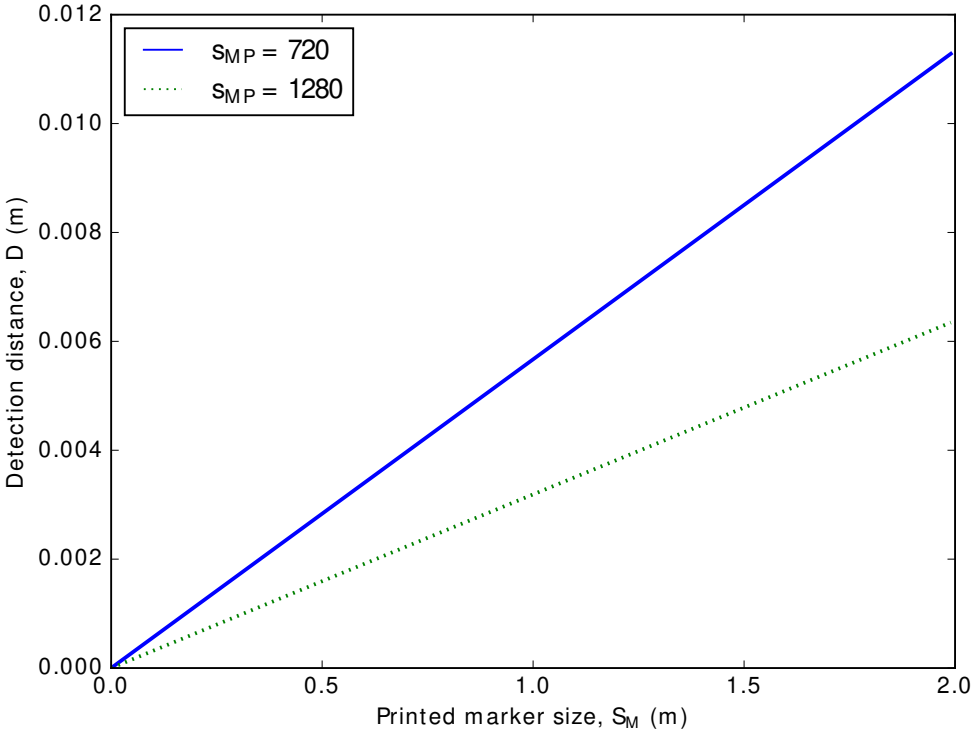


FIGURE 4.6: Variation of minimum detection distance, D, with printed markers.

4.3.1 Quantitative Assessment of Tracking Systems

The efficacy of marker-based and SLAM tracking systems was quantitatively evaluated using key performance metrics: **accuracy**, **precision**, **mean absolute error (MAE)**, and **standard deviation (SD) of error**. These metrics were essential for comparing the systems' reliability and adaptability in flight simulation training environments. Table 4.1 summarizes the main results.

These metrics were chosen to comprehensively evaluate the systems' effectiveness in enhancing flight simulation training. Accuracy and Precision directly impact the fidelity of virtual overlays, while scenario versatility and cognitive skills improvement reflect the systems' adaptability and potential for broader educational applications.

Accuracy indicates how close the measured positions (detected by the tracking system) are to the actual, true positions. It was calculated based on the proportion of correct detections, reflecting the system's ability to accurately locate and maintain the position of key objects within the simulation environment. The formula used for the accuracy rate is:

$$\text{Accuracy Rate (\%)} = \left(\frac{\text{Number of Correct Detections}}{\text{Total Number of Detections}} \right) \times 100 \tag{4.3}$$

This study determined accuracy by comparing the known, true positions (e.g.,

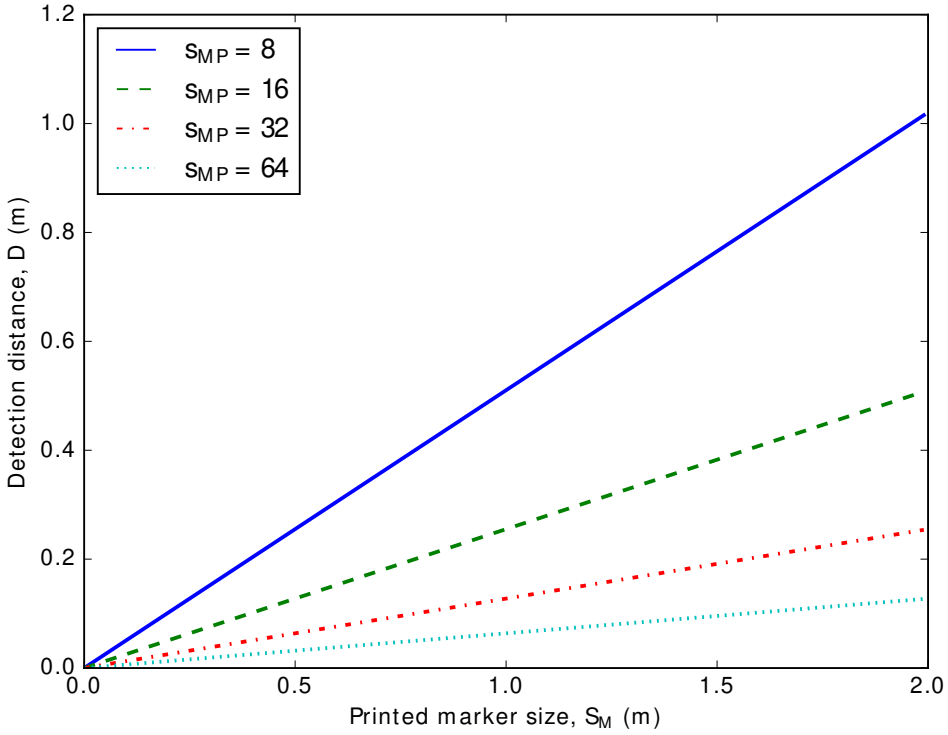


FIGURE 4.7: Variation of maximum detection distance, D, with printed markers.

coordinates set during calibration) with the detected positions obtained from the marker-based and SLAM systems. Multiple trials were conducted to ensure robustness and a correct detection was recorded if the system's output fell within a predefined acceptable error range (e.g., within a few millimetres).

Precision measures the consistency or repeatability of the tracking system. Even if a system is accurate, it might not be precise if it does not consistently detect the same position across multiple trials. The precision rate was calculated using the formula:

$$\text{Precision Rate (\%)} = \left(\frac{\text{Number of Consistent Detections}}{\text{Total Number of Repeated Measurements}} \right) \times 100 \qquad (4.4)$$

To assess Precision, the same position was repeatedly detected over multiple trials, and the variance among those measurements was evaluated. A higher consistency across readings translated to a higher precision rate.

To provide a detailed assessment of the tracking systems' performance, two key error metrics were computed: **Mean Absolute Error (MAE)** and **Standard Deviation (SD) of Error**.

Mean Absolute Error (MAE): This metric measures the average absolute difference between the true position (x_i) and the position detected by the tracking system ($\hat{x}_i$). It is a critical indicator of the overall accuracy of the system. The formula used is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{4.5}$$

Where:

- n is the total number of measurements,
- x_i is the true position, and
- $\hat{x}_i$ is the detected (predicted) position by the system.

Standard Deviation of Error (SD): This metric captures the variability or spread of errors, providing insight into the system's Precision. A lower standard deviation suggests that the system consistently detects positions close to the true value. The formula used is:

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4.6}$$

Where:

- $\bar{x}$ is the mean of the observed errors, calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i) \tag{4.7}$$

In addition to these error metrics, the **Scenario Versatility Index (SVI)** was introduced to evaluate the systems' ability to adapt across multiple training scenarios. SVI quantifies the number of distinct scenarios each system supports, with higher values indicating greater adaptability. For instance, a value of $SVI = 1$ represents functionality within a single scenario, while $SVI = 5$ indicates adaptability across five distinct scenarios. This metric underscores SLAM's capability to operate in dynamic and unstructured environments.

TABLE 4.1: Performance Metrics of Marker-Based and SLAM Systems

Metrics	Marker-Based Systems	SLAM Systems
Accuracy Rate (%)	98.5	96.2
Precision Rate (%)	97.8	94.5
Mean Absolute Error (cm)	0.5	1.2
Standard Deviation of Error (cm)	0.3	1.0
Localization Speed (Milliseconds)	16	30
Scenario Versatility Index (1 scenario)	1	1
Scenario Versatility Index (2 scenarios)	1	2
Scenario Versatility Index (3 scenarios)	1	3
Scenario Versatility Index (4 scenarios)	1	4
Scenario Versatility Index (5 scenarios)	1	5
Cognitive Skills Improvement Factor	4	7

4.3.2 Discussion

The high precision observed in marker-based tracking confirms the findings by Cheng et al. [176], who highlighted that marker-based augmented reality (AR) provides high positional accuracy, crucial for precise overlays in flight simulation. Our results validate this observation, as the marker-based system consistently demonstrated high accuracy and precision, making it ideal for tasks that require exact visual alignment. Cheng et al. also noted potential instabilities, such as shakiness, often linked to AR software development kits' marker quality and performance (SDKs). In our controlled simulation settings, these issues were effectively mitigated through proper calibration and stable environments, underscoring the importance of setup optimization in enhancing marker-based system reliability.

Conversely, our findings on SLAM align with Cheng et al.'s observations regarding the adaptability of markerless AR. By utilizing environmental features and sensor data (e.g., gyroscopes and accelerometers), SLAM systems offer flexibility, enabling the projection of virtual information onto diverse surfaces without relying on pre-placed markers. This adaptability enhances spatial awareness and cognitive skills during flight training scenarios despite the slightly lower accuracy observed than marker-based systems. The broader skill development facilitated by SLAM systems makes them particularly valuable in dynamic and unstructured training environments.

4.3.3 Impact of Sensor Resolution and Marker Size on Detection Range

The dependency of marker-based tracking systems on sensor resolution and marker size is evident from our results, reflecting similar findings by Rabbi et al. [177], who reported that increasing marker size significantly enhances the detection range. Our study extends this understanding by quantifying how variations in marker sizes influence detection thresholds at different distances.

As demonstrated in Figures. 4.6 and 4.7, the detection range for marker-based tracking systems is significantly affected by both sensor resolution and marker size. Larger markers not only extend the detection range but also improve the sharpness of their identification. This factor is critical for maintaining precision at extended distances, particularly in scenarios requiring precise alignment of virtual overlays with physical cockpit elements. Figure 4.8 further illustrates the relationship between marker size, sensor resolution, and detection range, offering insights into optimizing marker-based systems for AR applications.

The mathematical derivations used to calculate the detection range (see Eq. 4.2) emphasize the interplay between camera specifications—such as focal length (f), field of view angle (θ), and image size (s_{IP})—and marker size (S_M). These calculations highlight the importance of tailoring marker size and resolution settings to the specific requirements of the training environment.

While marker-based systems excel in controlled settings, SLAM systems offer distinct advantages in adaptability. Their ability to map and project virtual information onto any surface without pre-placed markers creates an immersive and dynamic experience. This makes SLAM ideal for training scenarios requiring real-time responsiveness and broader skill development, as depicted in Figure 4.9. By leveraging

FIGURE 4.8: Various stages of augmented reality marker integration in flight simulator training, showcasing different marker placements and sizes for enhanced cockpit interaction and tracking accuracy.

environmental features for tracking, SLAM systems provide flexible and engaging training experiences that enhance spatial awareness and other cognitive skills.

These findings highlight the trade-offs between marker-based and SLAM tracking systems. Marker-based systems excel in precision and reliability, particularly for tasks requiring exact alignment in controlled environments. However, their dependency on stable conditions and pre-placed markers limits their adaptability. On the other hand, SLAM systems offer unmatched flexibility and immersive experiences, making them ideal for dynamic training scenarios. Together, these technologies provide complementary strengths to be leveraged to design AR systems tailored to diverse flight simulation training needs.

4.3.4 Comparative Performance Metrics

Our analysis revealed a distinct advantage for each technology. Marker-based systems excel in focused skill development, where high precision is essential. This reliability leads to improved objective flight data, with long-term cost savings for targeted training due to minimal hardware investment once markers are set up. SLAM, on the other hand, fosters real-world, transferable skills through its immersive and adaptable nature. Although it may have higher initial costs due to computational hardware and software updates, SLAM presents a scalable solution applicable across diverse training scenarios. This comparative analysis underscores the critical trade-offs between precision and adaptability, directly addressing the study's objective of optimizing AR-based tracking systems for flight simulation training as shown in Table 4.2.

From a user experience perspective, marker-based tracking offers confidence in precision tasks, ensuring trainees can rely on stable visual aids. With their dynamic engagement, SLAM systems enhance cognitive skills and situational awareness, making them more suitable for training scenarios involving complex, adaptive tasks.

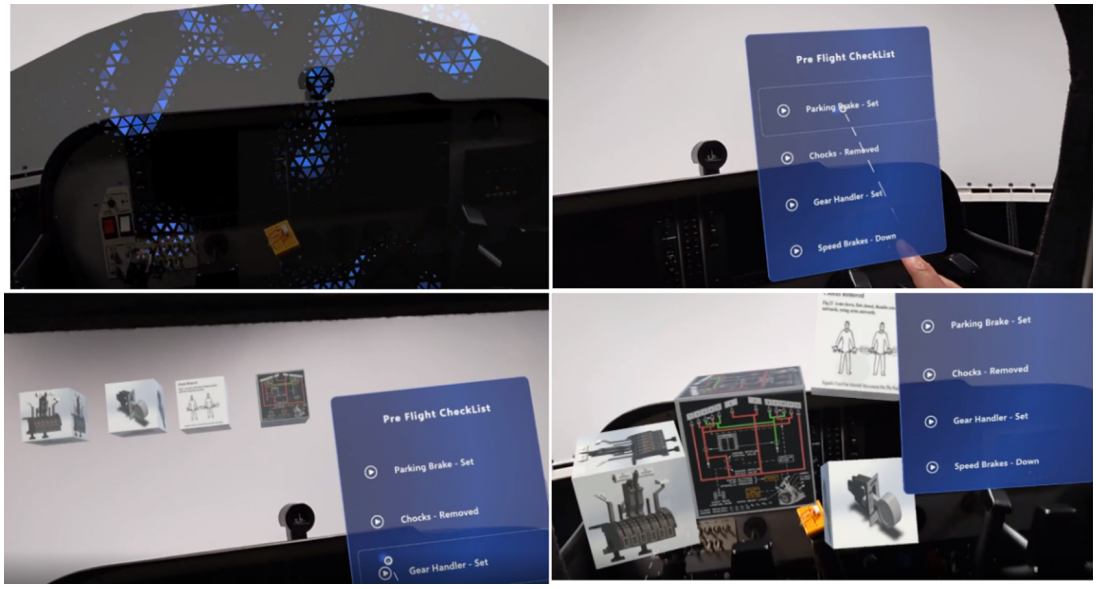


FIGURE 4.9: SLAM tracking in AR for procedural training: showcasing interactive checklists and equipment diagrams to boost pilot understanding and interaction.

TABLE 4.2: Comparison of Marker-Based and SLAM Systems for Simulation Training

Aspect	**Marker-Based Systems**	**SLAM Systems**
Accuracy and Precision	High precision in ideal conditions; accuracy drops if markers are occluded or not visible.	Comparable to markers in ideal conditions but can degrade in low-texture environments.
Stability	Very stable in controlled environments; issues arise when markers are occluded.	Adapts to environmental changes, offering a stable experience as the user moves.
Flexibility	Limited to environments where markers are maintained.	High flexibility, works in dynamic environments without predefined setup.
Scalability	Scalable within prepared environments.	Scalable to new environments without additional physical setup.
Cost-Effectiveness	Minimal hardware investment; costs increase with larger training areas.	May involve higher initial costs but provide long-term scalability and versatility.
Training Effectiveness	Ideal for training specific maneuvers.	Suitable for diverse training scenarios, enhancing cognitive skill transfer.

In our analysis, as depicted in Figure 4.10, the marker-based system reported an accuracy rate of 98.5% and a precision rate of 97.8%, with an MAE of 0.5 cm and an

SD of 0.3 cm. The SLAM system had an accuracy rate of 96.2% and a precision rate of 94.5%, with an MAE of 1.2 cm and an SD of 1.0 cm. These findings indicate that the marker-based system outperforms the SLAM system in terms of accuracy and precision, although the SLAM system offers greater flexibility in dynamic environments.

The derived error metrics underscore the trade-offs inherent to each system. Marker-based tracking provides higher reliability in controlled settings, as evidenced by its lower MAE and SD values. However, SLAM tracking's greater adaptability in changing conditions is shown by a broader error distribution and differences in localization speed. Marker-based systems boast rapid localization speeds of 16 milliseconds, indicative of their efficiency in stable environments. Although slightly slower, SLAM systems compensate with their adaptability, enabling real-time responsiveness in dynamic scenarios.

The versatility index in Figure 4.10 highlights the adaptability of the two systems. Marker-based systems excel at stable tasks, such as precise video overlays, while SLAM systems show a gradual increase in versatility across diverse scenarios. This performance, gauged by user accuracy and engagement, underscores SLAM's strength in dynamic environments.

Cognitive skills enhancement was also assessed. SLAM systems boosted cognitive abilities, scoring seven on a scale of 1 to 10, surpassing marker-based systems, which scored four. These measures were derived from objective data, such as response times and checklist accuracy, alongside users' self-evaluations of situational awareness.

The study reveals that while marker-based systems are best suited for controlled, precision-demanding tasks, SLAM systems are more adaptable and practical for dynamic, real-time training. These insights provide a foundation for designing hybrid AR systems that combine the precision of marker-based tracking with the adaptability of SLAM, offering a comprehensive solution for advanced flight training.

4.4 Conclusion

This chapter comprehensively analyses marker-based and SLAM tracking technologies within Microsoft HoloLens 2 for advanced flight simulation training. It highlights the unique strengths of marker-based systems in controlled environments, where their superior precision and reliability make them ideal for high-fidelity tasks that demand consistent overlay accuracy. These systems effectively reinforce trainees' comprehension of complex procedures by integrating interactive video tutorials and bridging virtual and physical interfaces in training environments.

Conversely, SLAM technology's adaptability and immersive capabilities make it a valuable asset for dynamic training scenarios. While it may involve higher initial costs, its scalability across varied training contexts and ability to support versatile and engaging experiences offer significant long-term benefits. This adaptability enhances spatial awareness and cognitive skills, making SLAM suited for complex, real-world training tasks.

This chapter synthesizes the comparative insights and directly addresses the thesis objectives of evaluating AR tracking technologies for flight simulation. It provides practical recommendations for developers and trainers, offering guidance on integrating these technologies effectively into training modules. Additionally, the analysis

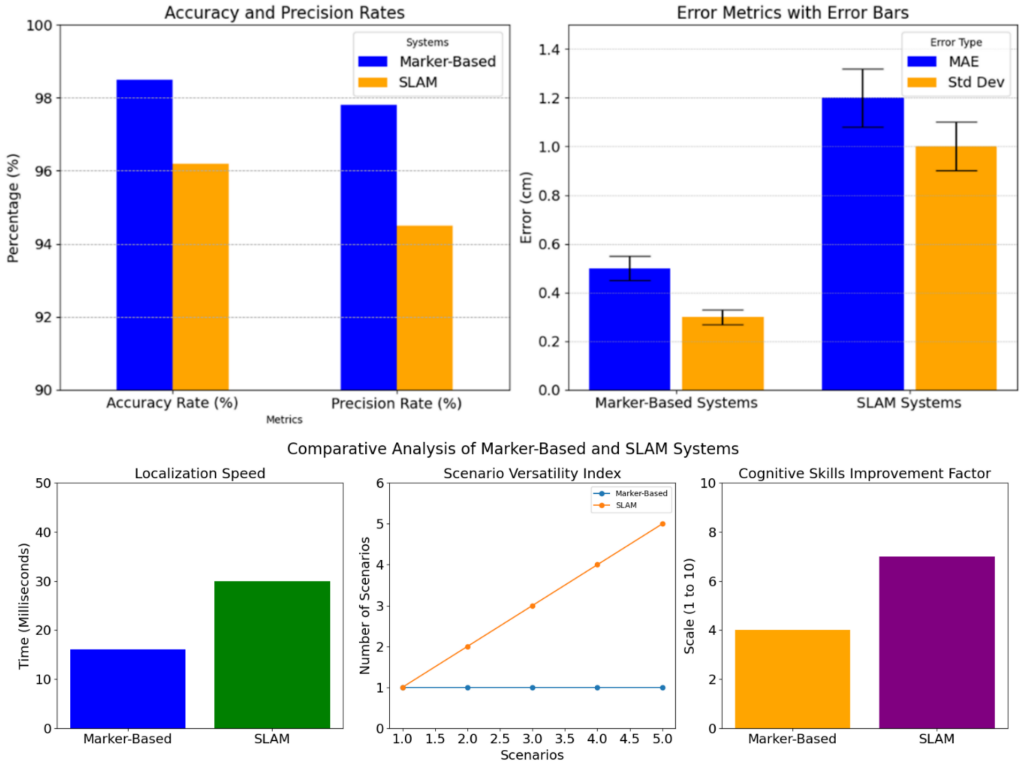


FIGURE 4.10: Comparison of marker-based and SLAM systems highlighting differences in accuracy, precision, error, localization speed, adaptability, and cognitive impact. These results emphasize the suitability of each system for specific training contexts.

identifies key trade-offs between precision and adaptability, serving as a foundation for advancing AR training systems.

Future development could explore a hybrid tracking approach that integrates the precision of marker-based systems with the flexibility and scalability of SLAM technology. Such a combination could further enhance the capabilities of AR devices like the HoloLens 2, addressing diverse training needs ranging from instrument-focused tasks to large-scale navigational exercises. Expanding this hybrid methodology holds promise not only for flight simulation but also for broader applications in medical training, automotive simulations, and military operations.

5 An Adaptation of Fitts' Law for Performance Evaluation and Optimization of Augmented Reality (AR) Interfaces

5.1 Introduction

Augmented reality (AR) transforms how users engage with digital information by seamlessly blending virtual elements into the physical environment [179], with applications spanning robotics, gaming, marketing, education, healthcare, simulation, safety, interior design, construction, agriculture, and quantum computing. This technological evolution is primarily driven by advancements in Simultaneous Localization and Mapping (SLAM), which integrates sensors like cameras and Inertial Measurement Units (IMUs) to ensure accurate alignment of digital elements within a user's real-world context [16]. SLAM enables AR experiences that are visually immersive and contextually relevant, enhancing user engagement by providing more interactive and responsive environments.

Despite rapid advancements, achieving intuitive and ergonomic user interactions in AR remains a significant challenge. For AR to reach its full potential in providing compelling user experiences, it is essential to develop interaction models that capitalize on natural gestures and body movements. Research demonstrates that well-designed user interactions in AR can enhance engagement, improve retention, and increase satisfaction—factors crucial for AR's adoption across diverse applications [180].

To bridge these challenges, this chapter addresses a pivotal gap in Human- Computer Interaction (HCI) by adapting Fitts' Law—a model traditionally applied to two-dimensional (2D) interfaces—to the complex three-dimensional (3D) environments characteristic of AR. Developed initially to predict movement time based on target distance and size, Fitts' Law has proven valuable in optimizing user interactions across many platforms [181]. However, extending its application to AR presents unique challenges due to the added spatial dimensions and dynamic orientation factors in 3D environments. In response, this study contributes a novel framework for modelling and predicting interaction complexities in AR by incorporating head movement and SLAM-based spatial awareness, addressing the distinct requirements of AR interfaces. Existing studies by [182], [183] have attempted to apply Fitts' Law to virtual 3D environments, but they often fall short when adapted to AR, where spatial

[0]Publication note: This chapter is based on the peer-reviewed article [178]. https://doi.org/10.1109/ACCESS.2024.3498444.

orientation and user expectations play critical roles. Misalignments between user expectations and interaction outcomes can decrease efficiency and user satisfaction. To address this, our study incorporates a comprehensive survey to capture subjective user feedback across different spatial orientations in AR and experimental data to validate the model's effectiveness. The study evaluates task difficulty across various AR spatial quadrants through this approach, applying Fitts' Law to assess and optimize interaction effectiveness in these environments.

Additionally, this research refines Fitts' Law parameters specifically for AR interfaces by employing a genetic optimization algorithm, enhancing the precision of AR interactions. This novel refinement approach allows for more responsive and user-centred AR experiences. Finally, we establish a benchmark for AR interaction models by combining insights from subjective surveys and empirical data. This ensures a robust understanding of AR interface performance and provides a reference for future AR studies.

These findings have substantial implications for AR development, advancing interaction modelling and usability in AR. They impact various applications, ranging from VR and AR gaming to complex simulation environments and interactive services, such as flight simulators. This work represents a significant step in optimising AR technology, equipping researchers and professionals with the insights needed to shape the future of user-friendly and efficient AR interfaces. In particular, this chapter directly supports the overarching objectives of this thesis by evaluating interaction modelling and tracking performance in AR, with a focus on their practical application to immersive flight simulation training using the Microsoft HoloLens 2.

5.1.1 Evolution of Fitt's Law

Fitts' Law has served as a foundational model in HCI. It was initially applied to Graphic User Interfaces (GUIs) to predict the difficulty of tasks involving object selection based on factors like target size, distance, and movement direction [184]. Over time, Fitts' Law has been adapted for various interaction modalities, including touchscreen interfaces [185], [186], [187] and Virtual Reality (VR) [188] environments, marking significant progress in HCI's exploration of interaction dynamics in different spatial contexts.

The evolution of Fitts' Law began with MacKenzie and Buxton's 1992 extension to 2D pointing tasks using a mouse [189], demonstrating that the model could effectively predict movement time for graphical user interface (GUI) interactions. Subsequent adaptations expanded its applicability to other contexts. For instance, Accot and Zhai's 2003 refinement for 2D interfaces introduced the "steering law," which modelled tasks requiring continuous movement through constrained paths, such as scrolling or navigating menus [190].

Similarly, Murata and Iwase's 2001 model for 3D environments accounted for spatial variables like azimuth angles, providing insights into how movement direction and spatial orientation influenced task difficulty in three-dimensional spaces [191]. Grossman and Balakrishnan (2004) further enhanced Fitts' Law by incorporating the effects of target size and movement direction in 3D environments, revealing that interaction efficiency depended on the alignment of the user's movement path with the spatial orientation of targets [192]. At the same time, Schuetz et al.'s studies on gaze-based interaction explored the application of Fitts' Law to eye-tracking technologies,

demonstrating that movement time predictions could be extended to dynamic scenarios where users relied on gaze direction rather than physical input devices. These findings highlighted the potential of Fitts' Law to model interaction effectiveness in increasingly complex and adaptive environments [193].

Recent studies have continued this trajectory, with research exploring Fitts' Law adaptations for VR and AR. For example, Jiang et al. (2020) [187] extensively reviewed Fitts' Law applications in 3D environments, and Wagner et al. (2023) [194] investigated gaze-hand alignment in 3D user interfaces. These studies highlight the adaptability of Fitts' Law across different modalities and its potential to model interaction effectiveness in immersive technologies.

Despite these developments, applying Fitts' Law to AR still needs to be explored, particularly in scenarios incorporating head movement and the spatial dynamics unique to AR. Current adaptations often need to adequately address AR's distinct interaction complexities, where factors like user orientation and environmental variability significantly impact task difficulty. Building on foundational studies in 2D and 3D environments, this research proposes an AR-specific model that integrates SLAM-based spatial awareness and head movement into Fitts' Law, contributing a refined approach that better addresses the nuances of AR interactions.

This chapter aims to bridge this gap by offering a model tailored to AR's intricacies, thereby advancing both the theoretical and practical understanding of user interactions in augmented reality.

5.2 Methodology

This research extends Fitts' Law to accommodate the specific characteristics of augmented reality (AR) by integrating head movement and optimizing critical parameters through a genetic algorithm tailored to these immersive environments.

5.2.1 Ergonomics of Interactive Element Placement in AR

The study applies Fitts' Law to examine user interactions across different sectors within a 360° AR environment. This framework enables a detailed analysis of the ease with which users can locate and interact with virtual objects positioned in their field of view, arranged according to cardinal and intercardinal directions, as represented in Figure 5.1.

Using Unity's XR Interaction Toolkit, virtual elements were precisely positioned within the AR environment, leveraging spherical coordinates to align with ergonomic principles. This ensures that objects are accessible and comfortable for user interaction, mainly by prioritizing target size and proximity in a three-dimensional space. Figure 5.2 illustrates the segmentation of the AR space into anatomical planes, providing a detailed analysis of interaction ease and difficulty across various spatial zones.

The analysis identified that users frequently engage with the upper frontal quadrant, suggesting ergonomic challenges. These findings align with prior studies emphasizing the importance of natural ergonomic zones to enhance usability and reduce strain during AR interactions [195].

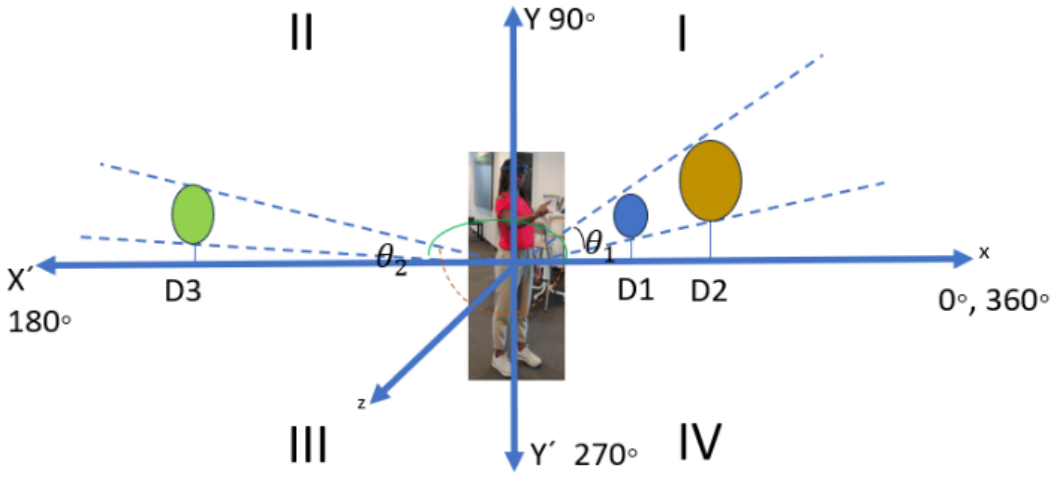


FIGURE 5.1: Framework for Ergonomic Placement and Interaction with AR Elements: This diagram illustrates the optimized positioning of virtual objects within a 360° AR environment, emphasizing ergonomic zones that support intuitive and comfortable user interactions.

5.2.2 SLAM-Based Spatial Anchoring in AR Object Placement

Building on the ergonomic framework, this study employs the simultaneous localization and mapping (SLAM) capabilities of the Microsoft HoloLens to dynamically position 3D virtual objects within an augmented reality (AR) environment. Algorithm 1 integrates real-time spatial awareness with user-centred design principles to ensure virtual objects' precise and context-aware placement, fostering a natural and immersive interaction experience.

The algorithm begins with initializing the SLAM framework, which generates a real-time spatial map of the user's surroundings. This spatial map forms the foundation for accurate object placement, allowing the system to adapt to environmental changes dynamically. The number of virtual objects to be instantiated in the AR scene is determined early in the process, with each object selected from a predefined set designed to offer diverse and engaging interaction scenarios. Using the Spatial Awareness system, the algorithm identifies optimal placement zones by considering the user's position, environmental obstacles, and ergonomic guidelines. These considerations ensure that virtual objects are within the user's visible and accessible range, avoiding occlusions or collisions with physical elements and adhering to ergonomic principles prioritizing comfortable interaction zones.

Object positions are initially computed in spherical coordinates relative to the user's location and converted into Cartesian coordinates using the HoloLens' spatial mapping capabilities. This transformation ensures that objects are precisely aligned within the user's field of view and maintains suitable interaction distances, enhancing usability and immersion. Additionally, variability in the scale and colour of each object is introduced during instantiation to enrich the interaction experience, promoting

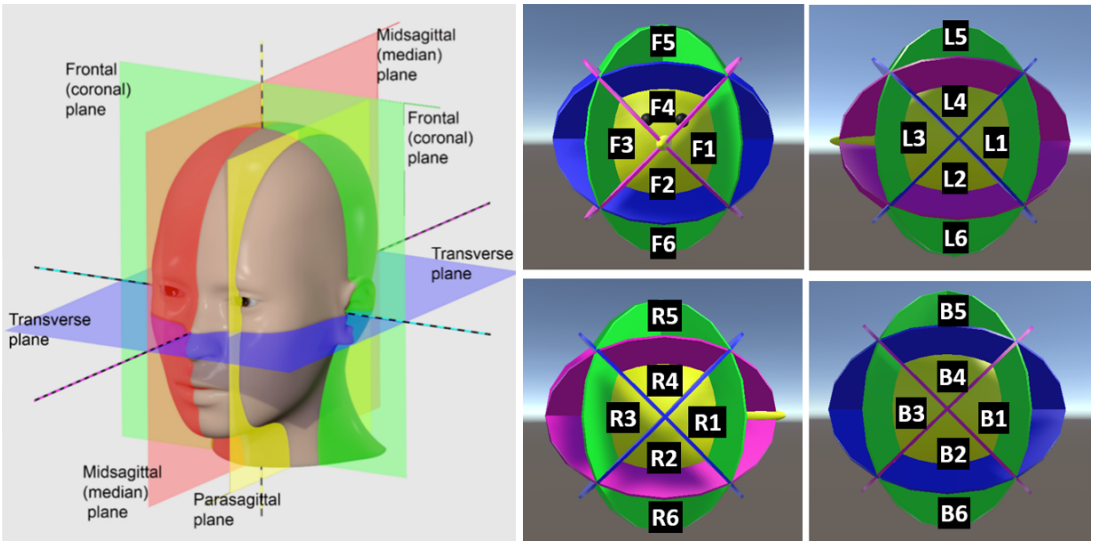


FIGURE 5.2: Anatomical Planes and Interaction Zones in AR Interface Design: This illustration compares the anatomical segmentation of the human body (left) [195] with the use of these planes in designing AR interaction zones (right), showing how quadrant-based zones can enhance intuitive interaction within AR interfaces.

user engagement and enabling diverse interaction scenarios. The proposed algorithm addresses key challenges in AR interactions by dynamically anchoring virtual objects in real-time while maintaining alignment with spatial and ergonomic guidelines. The system ensures intuitive positioning, enhances immersion, and supports usability in complex and dynamic environments.

Further details regarding the methodology and its implementation are available on the associated GitHub repository, providing a comprehensive resource for researchers and developers seeking to replicate or extend this work for various AR applications.

5.2.3 Adaptation of Fitts' Law for Augmented Reality Environments

In augmented reality (AR), task difficulty is influenced by numerous factors, necessitating refined metrics to evaluate user interactions effectively. This study leverages established and novel metrics to address the unique complexities of AR environments. Traditional Fitts' Law provides the **Index of Difficulty (ID)**, which quantifies interaction difficulty based on target size and distance. To account for the spatial and angular complexities inherent in AR, this research introduces the **Level of Difficulty (LoD)**. This novel metric adjusts for angular distances between the user and AR objects. Together, these metrics provide a comprehensive framework for assessing task complexity, with **Task Difficulty** as a general term encompassing both user-perceived and measurable difficulties during AR interactions. Throughout this work, precise references to specific metrics are made to maintain clarity and analytical rigour.

Algorithm 1 AR Object Placement and Interaction Data Collection

Require: C: CameraPosition ▷ The position of the AR camera in world space
Require: r: $[r_{\min}, r_{\max}]$ ▷ Range of distances for object placement
Require: O: [Sphere, Cube, Car] ▷ Array of 3D object prefabs
Ensure: L: [] ▷ List to store instantiated objects and their properties
InitializeCamera(C) ▷ Set up the AR camera at position C
$N \leftarrow$ DetermineSampleCount() ▷ Determine the total number of samples required
for $i \leftarrow 1$ to N **do**
    $T \leftarrow$ SelectRandomObject(O) ▷ Select an object type at random from O
    $[H, V] \leftarrow$ [SelectRandomHorizontalQuadrant(), SelectRandomVerticalQuadrant()] ▷ Select horizontal and vertical placement quadrants randomly
    $[d, \theta, \phi] \leftarrow$ [GenerateRandomDistance(r), CalculateAngle(H), CalculateAngle(V)] ▷ Generate spherical coordinates
    $[x, y, z] \leftarrow [d \cdot \sin(\phi) \cdot \cos(\theta), d \cdot \sin(\phi) \cdot \sin(\theta), d \cdot \cos(\phi)]$ ▷ Convert spherical coordinates to Cartesian coordinates
    $P \leftarrow C + [x, y, z]$ ▷ Calculate the object's world position relative to the camera
    SetObjectPosition(T, P) ▷ Set the object's position to P
    $[S, Col] \leftarrow$ [GenerateRandomScale(), GenerateRandomColor()] ▷ Generate random scale and color
    ApplyTransformations(T, S, Col) ▷ Apply the transformations to the object
    L.Add(T, CaptureMetrics(P, S, Col)) ▷ Add object and metrics to the list
    HandleUserInteraction(T) ▷ Handle user interaction with the object T
    $[Att, \Omega] \leftarrow$ CaptureInteractionMetrics(T) ▷ Record interaction time and orientation
    $N \leftarrow N - 1$ ▷ Decrement the sample count
    SaveData(L) ▷ Persist the collected data to storage
end for
if $N = 0$ **then**
    TerminateExperiment() ▷ Conclude the experiment and clean up resources
end ifreturn L ▷ Return the list of instantiated objects and their properties

Fitts' Law, as traditionally articulated, defines the index of difficulty ($\mathcal{I}_D$) using the following formula:

$$\mathcal{I}_D(S, T) = \log_2 \left[2 \frac{d(S, T)}{w(T)} \right] \tag{5.1}$$

Here, $\mathcal{I}_D(S, T)$ quantifies the interaction difficulty between a source object (S) and a target object (T), where $d(S, T)$ represents the Euclidean distance between the two objects, and $w(T)$ denotes the width of the target. The corresponding **Index of Performance (IP)** is defined as:

$$\mathcal{I}_P(S, T) = \frac{\mathcal{I}_D(S, T)}{\delta t(S, T)} \tag{5.2}$$

Where $\delta t(S, T)$ is the time required to move from the source to the target object. While these traditional formulations have been widely applied in 2D and static 3D

contexts, their direct application to AR environments needs to be improved due to the added complexities of spatial orientation and user movement.

In AR, interactions are not limited to static spatial parameters but often involve significant head or body rotation to align the user's gaze with the target. For example, tasks requiring head movement, referred to here as "with head movement," impose unique challenges compared to interactions where the user's head remains stationary, termed "without head movement." In these scenarios, traditional metrics like Euclidean distance and object width fail to capture the complexities of AR interactions fully. As a result, this study proposes an extended index of difficulty ($\mathcal{I}_{3D}$) that incorporates a head distance metric ($d_m(H,T)$) to account for head rotation and spatial positioning.

To calculate head distance and incorporate rotational factors into $\mathcal{I}_{3D}$, quaternion algebra is employed. Quaternion algebra provides a robust mathematical framework for representing and computing rotations in three-dimensional space. Specifically, the rotation of a vector p by a unit quaternion q is expressed as:

$$p' = q \cdot p \cdot q^{-1} \tag{5.3}$$

where $q^{-1} = q^*$ represents the conjugate of q. To quantify the rotational distance between two quaternions, q_1 and q_2, Świtoński et al. introduced the following formula:

$$d_Q(q_1, q_2) = \arccos(2\langle q_1, q_2 \rangle - 1) \tag{5.4}$$

This measure captures the angular difference between two orientations, accurately representing rotational complexity. Building on this approach, the present study incorporates quaternion-based head rotation calculations to model task difficulty in AR comprehensively.

Several essential modifications were made to the Fitts' Law for AR environments. First, the traditional width metric ($w(T)$) was replaced with the target object's volume ($V(T)$) to account for three-dimensional interactions. Additionally, the head distance metric ($d_m(H,T)$) was introduced to represent the spatial and angular relationship between the user's head (H) and the target object (T). This distance is critical for close proximity interactions, where targets positioned too close to the user may hinder visibility and interaction despite their reduced Euclidean distance.

The head distance metric reflects the relationship between task difficulty and spatial positioning. Task difficulty is hypothesized to be minimized at an optimal head distance (μ), representing the point where interaction is most efficient. Beyond this threshold, task difficulty increases as head distance deviates from μ. For example, objects placed too close to the user may require excessive head movement or distort visual perception, complicating the interaction. This relationship is illustrated in Figure 5.3, where the variation in task difficulty is plotted as a function of head distance.

The extended model also incorporates a genetic algorithm to optimize four critical parameters: Λ (scaling factor), μ (optimal head distance), σ (spread of head influence), and s (strength of head influence). These parameters allow the model to be tailored to the specific demands of AR interactions, ensuring that $\mathcal{I}_{3D}$ accurately reflects the complexities of spatial positioning and head movement in augmented reality.

This adaptation of Fitts' Law integrates quaternion-based head rotation, volumetric metrics, and optimized parameters, representing a significant advancement in modelling AR interactions. It provides a robust framework for evaluating task difficulty in AR environments, addressing the limitations of traditional models and offering new insights into the design and assessment of user interactions in immersive technologies.

5.2.4 Parameters in the Modified Fitts' Law

The task difficulty associated with head movement in augmented reality (AR) environments is modelled through the head distance metric ($d_m(H,T)$), mathematically expressed as:

$$d_m(H,T) = s - (s-1) \cdot e^{-\frac{1}{2}\left(\frac{d_Q(H,T)-\mu}{\sigma}\right)^2}$$

This formulation incorporates three critical parameters—s, μ, and σ—each of which plays a distinct role in quantifying the influence of head movement on task difficulty:

Strength of Head Influence (s): This parameter quantifies the impact of head movement on interaction task difficulty. It is derived from head movement data captured via HoloLens sensors, measuring the degree of head rotation required to align with each target object. A higher s value indicates a more substantial influence of head movement on task complexity. The parameter is optimized using a genetic algorithm, minimizing task completion time and ensuring efficient interactions.

Optimal Head Distance (μ): The parameter μ represents the ideal distance between the user's head and the target object, at which task difficulty is minimized. This value is empirically determined by analyzing task completion times across head-to-target distances. μ provides a benchmark for designing AR interfaces, ensuring optimal spatial alignment for user interactions.

Width of Head Distance Influence (σ): The parameter σ defines the range of distances over which head movement significantly affects task difficulty. Modelled as a Gaussian distribution centred around μ, σ quantifies the spread of head distance influence. This ensures the model accounts for near and far target placements within an AR environment.

These parameters—s, μ, and σ—are empirically determined through controlled experiments and optimized using a genetic algorithm to align the model with the unique requirements of AR interactions. Figure 5.3 illustrates the relationship between task difficulty and head distance, highlighting the roles of these parameters in shaping the interaction experience.

Building on these parameters, the study proposes a revised equation for the index of difficulty in 3D AR environments, incorporating head rotation and target object volume:

$$\mathcal{I}_{3D}(S,T) = \log_2\left[\Lambda \cdot d_m(H,T) \cdot \frac{d(S,T)}{V(T)}\right]$$

Here, $d(S,T)$ denotes the Euclidean distance between the source object (S) and the target object (T), while $V(T)$ represents the volume of the target. The parameter Λ serves as a scaling factor, ensuring the model's adaptability to diverse AR scenarios.

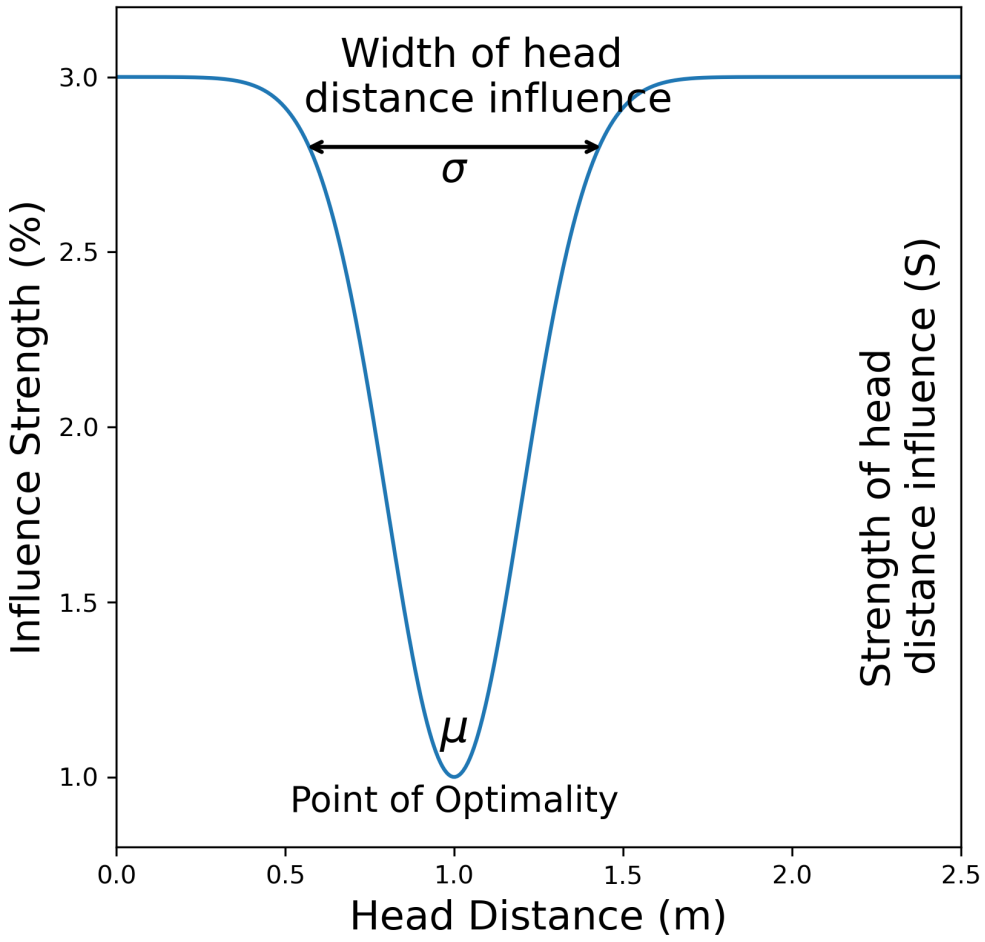


FIGURE 5.3: Illustration of the head distance influence with parameters s, μ, and σ.

Although the formula for the Index of Performance (IP) remains unchanged, it now integrates the revised index of difficulty ($\mathcal{I}_{3D}$), thereby enhancing its applicability to AR environments.

This adaptation of Fitts' Law offers a robust and refined metric for evaluating the complexities of AR interactions. By integrating head movement, spatial positioning, and object volume, the revised model provides valuable insights into interaction dynamics, particularly in immersive environments like those enabled by HoloLens. These advancements contribute to a more precise understanding of task difficulty in AR, facilitating the design of user-friendly and efficient AR systems.

5.2.5 Optimizing Fitts' Law Parameters for Augmented Reality

The experimental design of this study focuses on calibrating four pivotal coefficients in the revised Fitts' Law: Λ, μ, s, σ. These coefficients play a crucial role in adapting the model to the unique demands of augmented reality (AR) environments. A quantified Level of Difficulty (LoD) was introduced to provide a structured framework for evaluating task complexity, as illustrated in Figure 5.3. The LoD metric is derived from the normalized angular distance between the source and target objects, with normalization performed by dividing by π to confine the values within the interval $[0, 1]$. This normalization ensures consistency and interpretability across different experimental conditions.

The LoD values were discretized into four distinct categories to facilitate analysis:

$$LoD = 1 \Leftrightarrow 0.00 \leq \frac{d_m(H,T)}{\pi} < 0.25$$

$$LoD = 2 \Leftrightarrow 0.25 \leq \frac{d_m(H,T)}{\pi} < 0.50$$

$$LoD = 3 \Leftrightarrow 0.50 \leq \frac{d_m(H,T)}{\pi} < 0.75$$

$$LoD = 4 \Leftrightarrow 0.75 \leq \frac{d_m(H,T)}{\pi} \leq 1.00$$

Each experimental sample was assigned an initial LoD value, which subsequently informed the Index of Performance (IP) calculation. The IP was evaluated under two distinct conditions: with and without the influence of head distance. The samples were further divided into groups based on their assigned LoD values to gain deeper insights. For each group, the mean and standard deviation of LoD were calculated to assess the consistency and reliability of the categorization.

The range for the j-th group was defined as follows:

$$R(j) = \left[\text{avg}(LoD = j) - \text{std}(LoD = j), \text{avg}(LoD = j) + \text{std}(LoD = j)\right]$$

This approach ensured that the experimental framework could effectively capture the variations in task difficulty associated with different LoD levels. The calibration of the coefficients aimed to meet two critical conditions:

1. **Distinct Separation Between LoD Categories**: The average LoD values were required to exhibit a marked and increasing separation, specifically:

$$\text{avg}(LoD = 1) \ll \text{avg}(LoD = 2) \ll \text{avg}(LoD = 3) \ll \text{avg}(LoD = 4).$$

2. **Minimal Overlap Between LoD Ranges**: The ranges of consecutive LoD groups were required to have minimal or no overlap to ensure clear differentiation between levels of difficulty:

$$R(j) \cap R(j+1) = \varnothing \quad \text{or negligible.}$$

5.3 Experimrntal Setup

The experimental design guarantees that the LoD categories are robustly defined and effectively capture the gradations of task difficulty by ensuring distinct separation and minimal overlap. This structured calibration approach enhances the applicability and precision of the revised Fitts' Law model in AR environments. It supports its use in designing user-centred AR systems that account for varying interaction complexities.

Experimental Procedure: This chapter details the experimental setup employed in this study, leveraging Microsoft HoloLens 2, a state-of-the-art augmented reality (AR) device. The selection of the HoloLens 2 was guided by its sophisticated sensor suite, particularly its advanced head-tracking capabilities, which are critical for evaluating the impact of head movement on task difficulty during interactions in augmented reality environments.

Participant Recruitment and Selection: A total of 30 participants, including twenty-four males and six females, aged between 18 and 50 years (mean = 32, SD = 13.39), were recruited for the study. Recruitment was conducted through a multi-channel approach to ensure a diverse and representative sample across age, gender, and prior AR experience. This approach facilitated the collection of interaction data that reflected varied user profiles and behaviors within AR environments. Prior to participation, individuals were briefed on the scope and nature of the study, ensuring a clear understanding of the objectives, procedures, and any potential risks. This briefing also emphasized the voluntary nature of participation, with all participants informed of their right to withdraw at any time without consequence. Confidentiality of participant information was assured throughout the study.

Task Description: Participants were assigned to one of two groups based on posture: standing or seated, as depicted in Figure 5.4. Each participant wore the Microsoft HoloLens 2 headset and was immersed in a controlled AR environment to facilitate interactive tasks. The experimental protocol began with participants completing a registration process where they entered their username, session ID, and the number of samples they intended to interact with during their session. The duration of each session ranged from 10 to 15 minutes, depending on the number of samples selected. Participants with more significant experience in AR environments typically opted to interact with a more significant number of samples (40-60), while those with limited experience chose fewer (10-30). To ensure consistency in data collection and maintain participant focus, no breaks were provided during the sessions.

The interactive tasks commenced with a guiding arrow appearing in the participant's field of view, directing them to a holographic target. These targets were strategically placed at varying distances, ranging from 0.3 to 6.0 meters, and distributed across different spatial quadrants within the AR environment. This arrangement was designed to introduce spatial variation and encourage diverse head movements. The holographic objects varied in size, ranging from 0.1 to 1.2 meters, with scaling factors applied between 0.3 and 0.5 to ensure a heterogeneous interaction experience. This variability allowed participants to engage with objects of different dimensions and distances, replicating various interaction scenarios.

To maintain task continuity, the orientation of the guiding arrow was updated after each interaction to indicate the next target. A standardized two-second intermission was implemented between interactions, accompanied by a brief prompt to prepare

participants for the subsequent task. The objects encountered during the session included spheres, cubes, and cars, which were selected to analyze the impact of object form and participant posture on interaction modalities and user experience. The array of tasks and object variations ensured a comprehensive examination of the factors influencing interaction within the AR environment. This setup provided the data for evaluating performance metrics and understanding user dynamics in different ergonomic and spatial conditions.

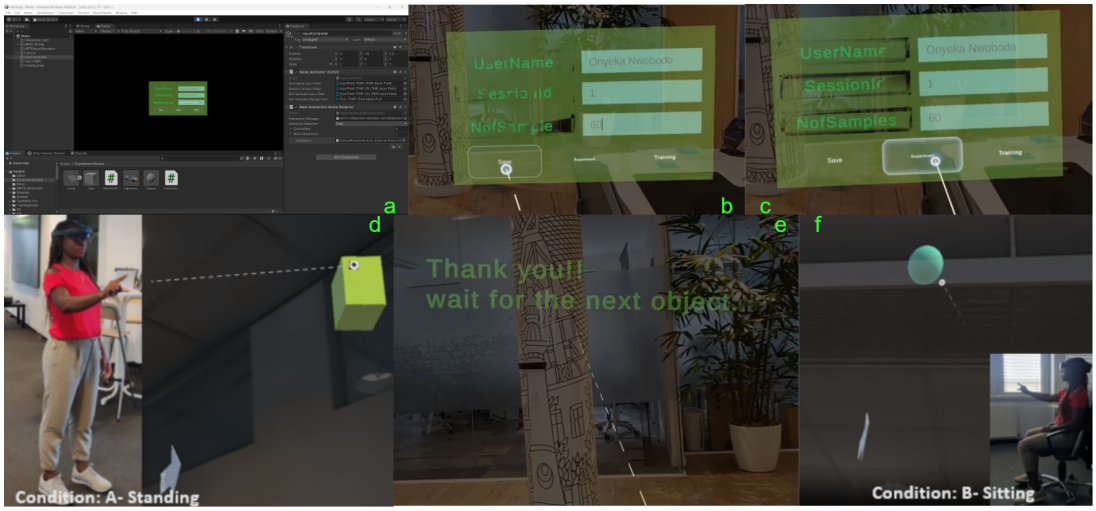


FIGURE 5.4: Experiment setup: A user interacts with a virtual environment through a HoloLens device, performing tasks in both standing (Condition A panel(d)) and sitting (Condition B panel(f)) positions. The setup includes interface development in Unity panel(a), user input registration panel(b,c), and interaction with virtual objects panel (e).

5.3.1 Data Collection and Dataset Overview

The dataset used in this study was meticulously recorded using the software tools described earlier, ensuring precise and reliable data collection. Each data entry captures several critical parameters related to the interaction, enabling a comprehensive analysis of user behaviour and performance within the AR environment.

Data Parameters Each dataset entry includes the following components:

Object Location and Observer Head Orientation: The dataset records the position and rotation of the object within the AR environment alongside the observer's head orientation. This data provides a detailed understanding of spatial relationships and movement dynamics during interactions.

Distortion Vector and Scaling Factor: The distortion vector quantifies the object's scaling across various axes. This vector adheres to a predefined range of $[0.8, 1.2]$, ensuring the object's geometric integrity remains intact throughout the interaction. The scaling factor derived from the distortion vector is used to calculate the object's volume, providing insights into the spatial dimensions experienced by the participant.

Action Time (δt) Action time is defined as the interval between the initiation of an application and the participant's interaction with the object. This metric is critical

for computing the Index of Performance (referenced in Equation 5.1). A standardized two-second interval adjusts the time difference between successive interactions to account for experimental consistency and ensures reliable comparisons across tasks.

Data Collection Framework The data collection framework was designed to systematically capture interaction details that enable nuanced analyses of user performance. This detailed framework allows for examining crucial interaction parameters, including object size, placement, participant posture, and the time required for task completion. The dataset supports a robust investigation into the dynamics of user interactions within the AR environment.

Dataset Composition: The dataset comprises 659 samples, with 438 recorded for standing and 221 for seated conditions. These samples provide a balanced representation of interaction scenarios across the two postures. The division of the dataset into Levels of Detail (LoD) is summarized in Table 5.1, highlighting the distribution of tasks and the corresponding parameters.

This comprehensive dataset forms the foundation of the experimental investigation, enabling a detailed exploration of how spatial dynamics, object variability, and posture influence interaction performance within AR environments. The rigour and precision of the data collection process ensure that this study's findings are grounded in robust empirical evidence.

TABLE 5.1: Data set characteristics. The number of samples and division into sitting and standing cases (horizontally) and into the Level of Difficulty (vertically).

No	Conditions	LoD=1	LoD=2	LoD=3	LoD=4	Total
1	Sitting	95	60	46	20	221
2	Standing	150	108	92	88	438
	Total	245	168	138	108	659

5.4 Results

This chapter presents and discusses the study's findings, highlighting the proposed model's effectiveness in improving predictive accuracy for augmented reality (AR) interactions and reducing interaction task difficulty. The results are substantiated through quantitative analysis and empirical validation, emphasizing the significance of spatial and ergonomic factors in AR environments.

Predictive Accuracy of the Refined Model: The refined model demonstrates a significant improvement in the predictive accuracy of AR interactions. This is evidenced by a strong correlation ($R^2 = 0.910$) between the predicted and actual user performance data, as illustrated in Figure 5.12. The reliability of the model is further confirmed through a QQ plot (Figure 5.13), which validates the normal distribution of residuals. These findings underscore the model's predictive power across diverse interaction scenarios.

Optimization of Interaction Task Difficulty: A genetic optimization algorithm was employed to fine-tune key parameters of the modified Fitts' Law (I3D) to enhance

the model's applicability. These parameters include: μ**:** Optimal head distance, σ**:** Width of head distance influence, s**:** Strength of head influence.

The optimization process resulted in a 40% reduction in interaction task difficulty across various spatial quadrants. This reduction was measured against the baseline values calculated using the modified Fitts' Law before optimization. The improvement was particularly pronounced in the upper frontal quadrant, a region with significant ergonomic challenges. The results of this optimization process were validated using empirical data from a flight simulation scenario, which closely replicates real-world AR applications.

The consistency of these findings across different spatial orientations highlights the proposed model's robustness in predicting and mitigating AR interaction challenges.

Influence of Spatial Positioning on Task Difficulty: The spatial positioning of objects within the AR environment significantly impacts interaction task difficulty. Objects within the user's direct line of sight (0° to 90°) are more accessible, whereas those positioned outside this range demand more complex head and body movements. This variation in accessibility underscores the importance of spatial ergonomics in AR interface design.

Tables 5.2 and 5.3 provide a detailed analysis of interaction task difficulty understanding and sitting conditions, respectively. At the same time, Figure 5.7 visualizes the average and standard deviation of the Index of Difficulty (ID) across various conditions. These results offer insights into how posture and spatial orientation influence user performance.

Analysis of Absolute Errors by Quadrant: The distribution of absolute errors across the four quadrants was examined using understanding and sitting conditions. Figure 5.8 highlights these findings:

Standing Condition: Absolute errors exhibit higher variability in the first and fourth quadrants (Q1 and Q4). This indicates increased task difficulty in these regions, likely due to the ergonomic challenges of reaching or viewing objects outside the user's direct line of sight.

Sitting Condition: The distribution of absolute errors across all quadrants is more consistent, with reduced variability compared to the standing condition. This consistency suggests that sitting provides users with more excellent stability and minimizes physical strain, improving interaction performance across quadrants.

These findings emphasize the critical role of spatial orientation and posture in AR interactions. The segmentation of results by quadrant demonstrates that ergonomic and positional factors significantly influence task difficulty. Notably:

- The upper frontal quadrant presents the most significant challenge, necessitating careful consideration in AR interface design.
- Sitting reduces variability in interaction difficulty, suggesting that posture stability is crucial in mitigating spatial challenges.

5.4.1 Statistical Analysis

This section presents the results of the statistical analyses conducted to examine the factors influencing task difficulty in augmented reality (AR) interactions. The analyses

include the effects of target location, lighting conditions, and environmental complexity on task performance under both standing and sitting conditions.

Target Location and Task Difficulty

Repeated-measures ANOVAs were performed with Tukey HSD post hoc tests at the 5% significance level to assess the impact of target location on task difficulty. The normality of residuals was confirmed using QQ plots (Figure 5.13), which demonstrated that the residuals followed a normal distribution.

The ANOVA results revealed a significant main effect of target location on task difficulty for both standing and sitting conditions ($F(3, N) = 46.51, p < 0.001$). Post hoc analyses identified significant differences between Quadrant 1 (Q1), Quadrant 2 (Q2), and Quadrant 4 (Q4) for the standing condition and between Q1 and Quadrant 3 (Q3) for the sitting condition. These differences are marked with asterisks in Figure 5.7.

Despite the observed differences in task difficulty, an analysis of the performance variability (STD) revealed no significant differences in error variability between quadrants for either condition ($F(3, N) = 1.53, p = 0.257$), as shown in Figure 5.8. This finding indicates that task performance was consistent across all quadrants.

Post-Optimization Analysis: Following the application of a genetic optimization algorithm, a significant main effect was observed for both standing and sitting conditions ($F(3, N) = 148.99, p < 0.001$). Post hoc analyses highlighted significant reductions in task difficulty for Q1 and Q4 under the standing condition and for Q1 and Q3 under the sitting condition. The optimized results are visualized in Figure 5.15, emphasizing the efficacy of the optimization in mitigating task difficulty.

Lighting Condition Analysis

The influence of lighting conditions on task difficulty was evaluated across spatial quadrants under both standing and sitting conditions. Lighting was categorized into three levels: low, medium, and high brightness. The Index of Difficulty (IoD) was analyzed using ANOVA and Tukey HSD post hoc tests.

The ANOVA revealed a significant effect of brightness on task performance ($F = 4.16, p = 0.016$), indicating that variations in brightness levels significantly influenced task difficulty. Post hoc tests identified Quadrant 4 (Q4) as the most affected by high brightness levels, which resulted in increased task difficulty (F-statistic: 3.37, p-value: 0.038). These findings are illustrated in Figure 5.5, where asterisks mark the significant impact in Q4. This suggests that visual strain and ergonomic factors may contribute to heightened task difficulty in this quadrant.

In contrast, no significant effect of brightness on task difficulty was observed for the sitting condition ($F = 0.581, p = 0.560$). Task difficulty remained consistent across all brightness levels in the sitting condition, as depicted in Figure 5.5. These findings highlight that standing tasks, particularly in Q4, are more sensitive to lighting conditions, whereas sitting tasks are relatively unaffected.

Environmental Complexity in AR

To examine the impact of environmental complexity on task performance, two factors were analyzed: object crowding (object count within quadrants) and object positioning (horizontal and vertical). Interaction time was used to measure task difficulty, and ANOVA tests were conducted to assess these factors.

Crowding: Objects were categorized into low, medium, and high levels of crowding based on their density within spatial quadrants. The ANOVA did not reveal a significant effect of crowding on task difficulty ($F = 3.52, p = 0.0602$). This suggests that the number of objects within a quadrant did not substantially influence interaction time or task difficulty.

Object Positioning: The analysis of object positioning revealed that horizontal placement significantly impacted task difficulty ($F = 10.73, p = 0.0025$), while vertical placement did not ($F = 1.56, p = 0.2658$). Post hoc Tukey HSD tests identified significant differences in task difficulty between Horizontal Quadrant 2 (HQuadrant 2) and Horizontal Quadrant 4 (HQuadrant 4) ($p = 0.024$). Figure 5.6 illustrates these effects, emphasizing the role of horizontal positioning in driving task performance challenges.

TABLE 5.2: The Average (AVG) and Standard Deviation (STD) of Task Difficulty Measures in Standing Conditions (Horizontally and Vertically) without and with head influence.

	Q1		Q2		Q3		Q4	
	HLoD	VLoD	HLoD	VLoD	HLoD	VLoD	HLoD	VLoD
Without Head Influence								
AVG	1.122	1.098	0.911	0.918	0.951	0.927	1.112	1.200
STD	0.733	0.695	0.566	0.633	0.685	0.652	0.856	0.897
With Head Influence								
AVG	1.534	1.503	1.264	1.255	1.288	1.339	1.512	1.643
STD	0.945	0.906	0.769	0.835	0.912	0.912	1.089	1.162

TABLE 5.3: The Average (AVG) and Standard Deviation (STD) of Task Difficulty Measures in Sitting Conditions (Horizontally and Vertically) without and with head influence.

	Q1		Q2		Q3		Q4	
	HLoD	VLoD	HLoD	VLoD	HLoD	VLoD	HLoD	VLoD
Without Head Influence								
AVG	1.971	1.765	1.450	1.649	1.955	2.019	2.224	2.094
STD	1.329	1.110	0.844	0.948	1.133	1.203	1.370	1.414
With Head Influence								
AVG	2.337	2.068	1.720	1.946	2.296	2.369	2.620	2.494
STD	1.581	1.327	1.003	1.130	1.382	1.407	1.609	1.707

These findings underscore the critical importance of optimizing lighting conditions and spatial organization in AR environments. The results indicate that:

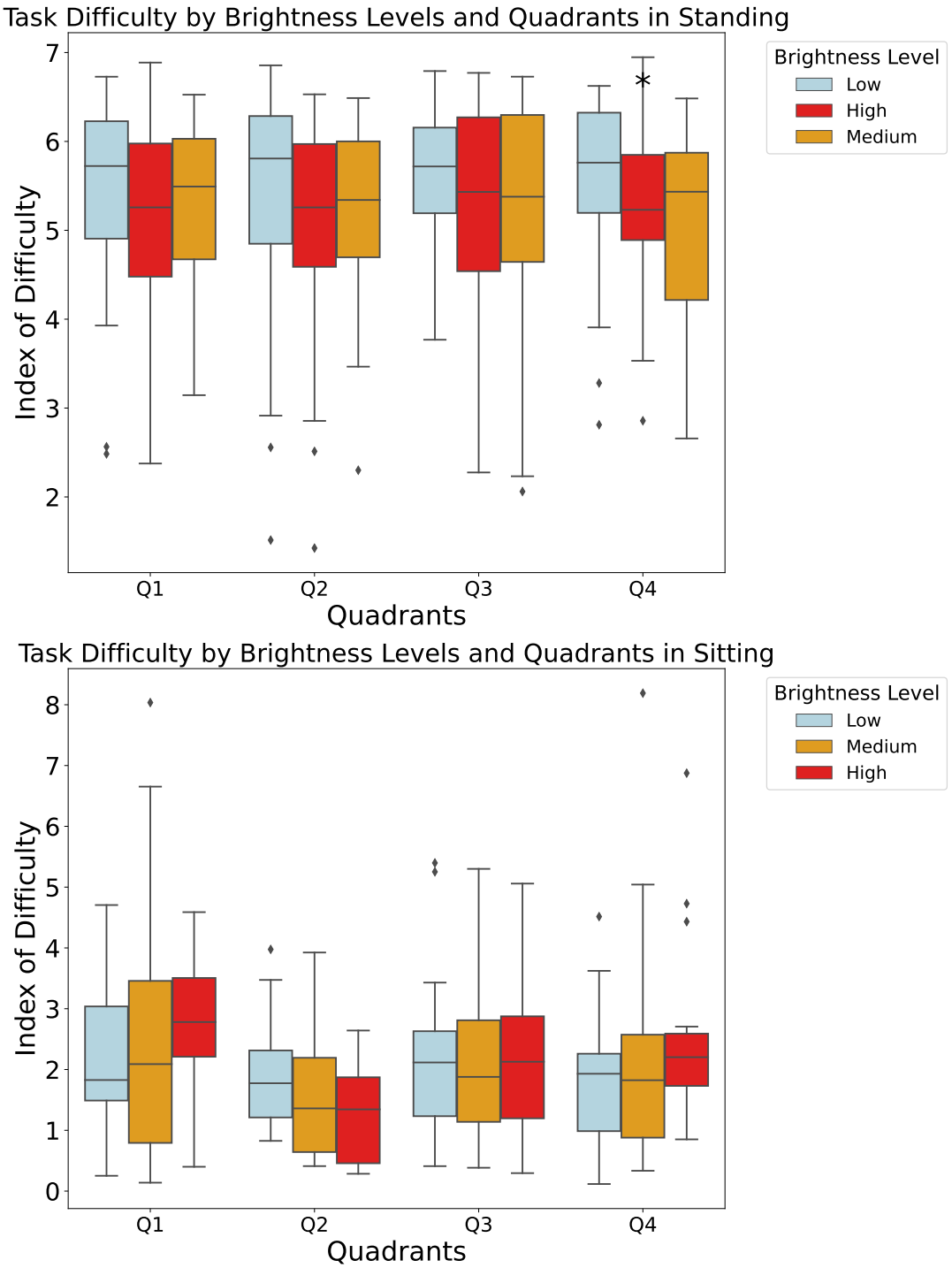


FIGURE 5.5: The Figure compares the Index of Difficulty (IoD) across four quadrants (Q1 to Q4) under three brightness levels (Low, Medium, and High) for both standing (top) and sitting (bottom) conditions. The asterisks (*) in the standing condition indicate significant differences in task difficulty ($p < 0.05$), with Quadrant 4 showing the greatest sensitivity to high brightness levels.

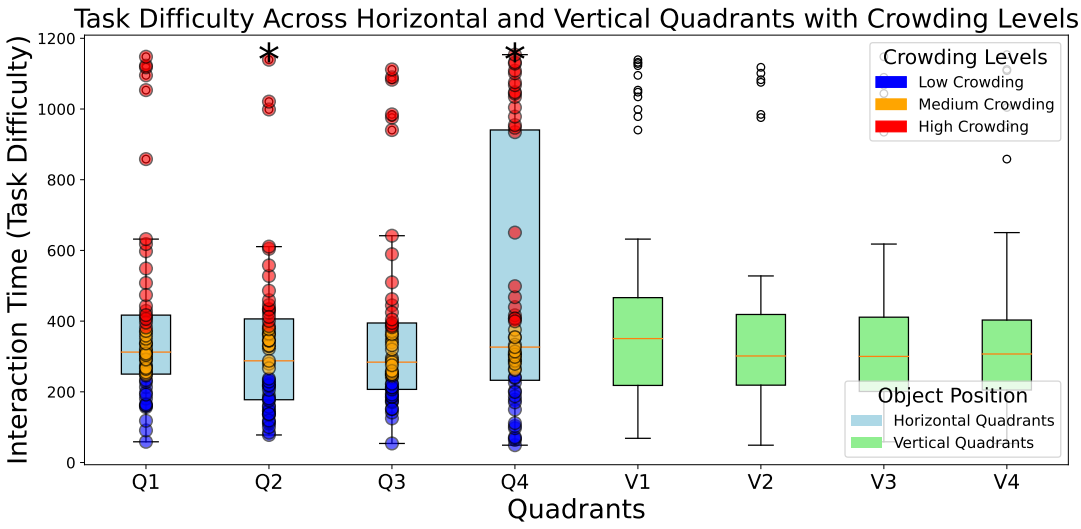


FIGURE 5.6: Boxplots show task difficulty (interaction time) across horizontal (Q1–Q4) and vertical (V1–V4) quadrants under different crowding levels (Low, Medium, High). Object positioning includes both horizontal and vertical quadrants. Asterisks (*) indicate significant differences in task difficulty for horizontal quadrants ($p < 0.05$), particularly in Q2 and Q4.

- Standing tasks are more sensitive to lighting conditions, particularly in Quadrant 4, suggesting that visual strain or ergonomic constraints may contribute to heightened task difficulty.
- Sitting tasks are less affected by brightness variations, with task difficulty remaining consistent across lighting levels.
- Horizontal positioning plays a significant role in determining task performance, while crowding and vertical positioning have a lesser impact.

By addressing these factors, AR system designers can reduce visual fatigue and improve user performance, particularly for standing tasks in challenging spatial orientations. These insights provide a foundation for developing ergonomic and efficient AR environments.

We further integrate horizontal and vertical dimensions into a unified framework to deepen the analysis of user task difficulty in augmented reality (AR) environments.

Multidimensional Approach to Task Difficulty: The three-dimensional nature of AR environments necessitates a holistic approach to evaluating task difficulty. Our research introduces a comprehensive metric that combines horizontal and vertical dimensions using the cosine and sine of user movement angles as primary indicators. The quadrant-specific behaviour of cosine and sine functions forms the foundation of this methodology. Table 5.4 defines the angle ranges and conditions for determining quadrant-specific interactions. This approach is particularly relevant for spatial navigation, where horizontal and vertical axes influence user movements.

Formulation of Task Difficulty Metric: To quantify task difficulty in AR, we propose a metric D_Q for each quadrant Q, expressed as:

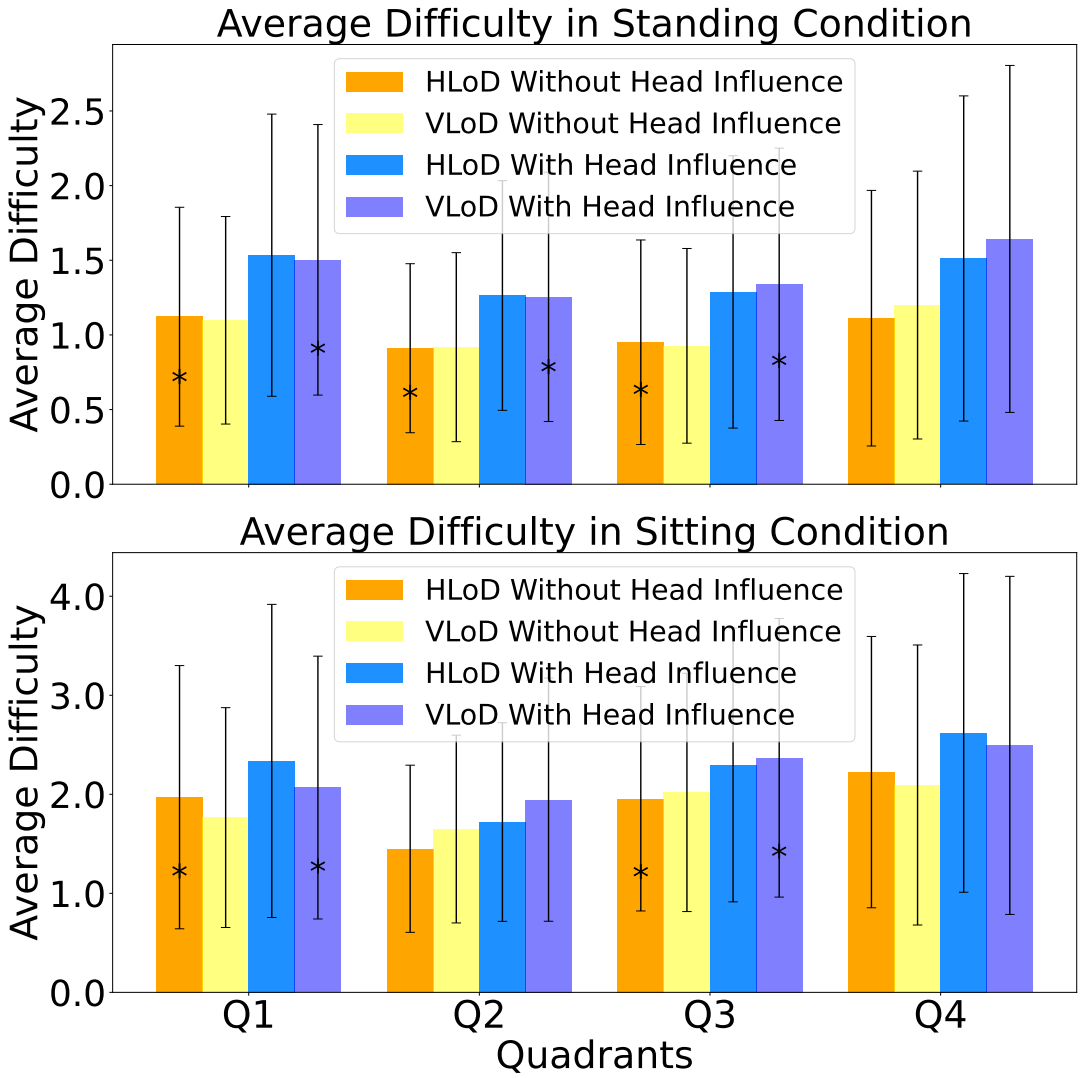


FIGURE 5.7: The bar charts show the average perceived task difficulty of AR interactions while standing and sitting across four quadrants (Q1-Q4), both with and without head movement. The asterisks (*) indicate quadrants with statistically significant differences ($p < 0.05$). These charts highlight increased task difficulty with head movement, emphasizing the importance of ergonomic design in AR interfaces.

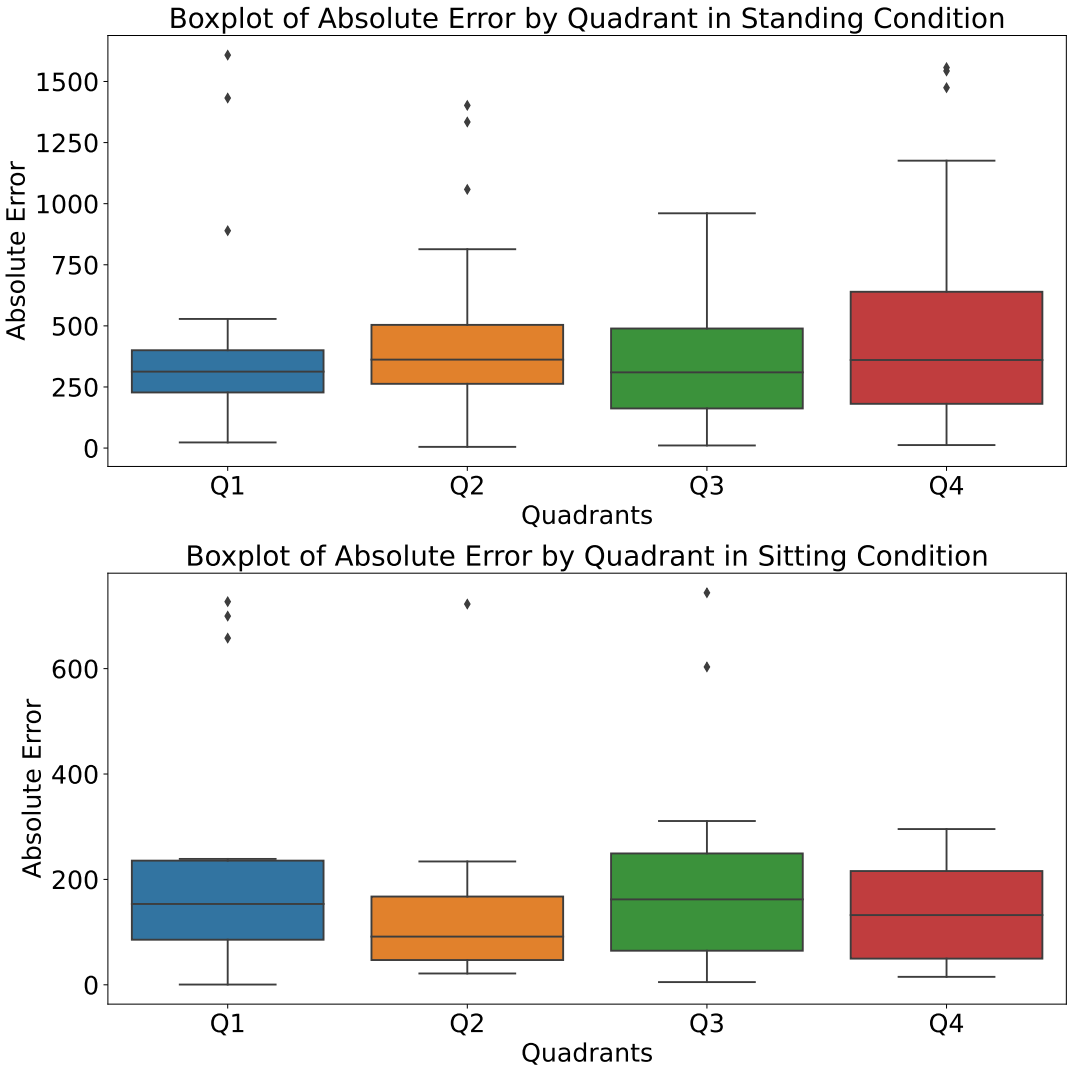


FIGURE 5.8: This figure shows the absolute errors across four quadrants (Q1 to Q4) during AR interactions in standing (top) and sitting (bottom) conditions. Each boxplot displays the error variability, with outliers marked as points, highlighting the precision and consistency of interactions across the conditions.

$$D_Q = \sqrt{(\cos_{\text{AVG}})^2 + (\sin_{\text{AVG}})^2} \tag{5.5}$$

This equation employs the 2D Euclidean norm, aggregating squared cosine and sine values to represent user movement's horizontal and vertical components. These components correspond to the relative task difficulty across different quadrants, as detailed in Table 5.4.

Including trigonometric components enables precise mapping of task difficulty to specific quadrants, capturing the directional challenges users face. This metric is crucial for evaluating user interactions, particularly in spatially complex AR environments.

Normalization for Consistency in 3D Environments: Recognizing the three-dimensional nature of AR interactions, the task difficulty metric was adapted for 3D contexts through normalization. The metric was scaled to a 0–1 range by dividing by the maximum value of $\sqrt{2}$, the upper bound of the Euclidean norm in 2D space when both horizontal and vertical components contribute equally. This normalization ensures consistency across spatial dimensions, minimizing bias from quadrant-specific complexities:

$$D_{Q_{\text{norm}}} = \frac{D_Q}{\sqrt{2}} \tag{5.6}$$

This normalization process ensures that interaction task difficulty is represented consistently, offering an unbiased evaluation across different spatial orientations. The resulting normalized metric provides a precise, standardized measure of task difficulty.

Integrated Visualization of Task Difficulty: The proposed methodology's effectiveness is demonstrated through heatmaps (Figure 5.9). These heatmaps compare average task difficulty in standing and sitting positions, with and without head movement. The colour gradients visually represent the difficulty level, with lighter shades indicating higher challenges.

The heatmaps reveal that head movement significantly increases task difficulty, particularly in standing positions. This result highlights the significance of considering head movement in the ergonomic design of AR systems. Integrating horizontal and vertical dimensions into a unified metric offers a transparent and standardized method for evaluating user challenges, as shown in Equation 5.5.

5.4.2 Quantitative Evaluation of User Interaction Challenges

To evaluate the perceived task difficulty associated with interacting with augmented reality (AR) targets, we employed a **Weighted Average Difficulty (WAD)** metric derived from user ratings. This metric, summarized in Table 5.5 and illustrated in Figure 5.11, offers a proportional representation of perceived task difficulty across various user interactions. The WAD metric was calculated using the following formula:

$$WA = \sum_{i=1}^{n} (P_i \times V_i) \tag{5.7}$$

Here, WA represents the weighted average difficulty, P_i corresponds to the percentage of responses for each difficulty level i, and V_i denotes the ordinal value assigned to the difficulty level (ranging from 1 to 7, from "Very Easy" to "Extremely Hard"). This

methodology ensures that each level of perceived difficulty is proportionally reflected in the final calculation, accurately representing the collective user experience.

Trigonometric conditions were used to classify user interactions across spatial quadrants to enhance the precision of our analysis. These conditions are detailed in Table 5.4, which defines the angle ranges for determining the quadrant corresponding to specific interactions. The quadrant classification leverages the behaviour of cosine and sine functions to categorize user movements into specific spatial orientations.

TABLE 5.4: Trigonometric Conditions for Quadrant Determination

Condition	Range of y
If $\cos y > \frac{1}{\sqrt{2}}$ and $\sin y \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$	$\left[0, \frac{\pi}{4}\right]$
If $\sin y > \frac{1}{\sqrt{2}}$ and $\cos y \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$	$\left[\frac{\pi}{4}, \frac{3\pi}{4}\right]$
If $\cos y < -\frac{1}{\sqrt{2}}$ and $\sin y \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$	$\left[\frac{3\pi}{4}, \frac{5\pi}{4}\right]$
If $\sin y < -\frac{1}{\sqrt{2}}$ and $\cos y \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$	$\left[\frac{5\pi}{4}, \frac{7\pi}{4}\right]$

Our analysis of WAD values across various task series revealed significant patterns and trends: The F Series exhibited moderate variance in difficulty, ranging from 2.38 to 4.35, with task F4 being particularly challenging. The L Series showed a narrower range of difficulty, peaking at 5.29 for task L6. The R Series was consistently challenging, with the highest difficulty recorded for task R6 (5.36). The B Series demonstrated the broadest range of difficulty, with B6 being the most demanding task at 5.50.

Interestingly, specific tasks shared identical WAD values, such as L4 and R3/R4 (3.34 and 3.71, respectively). The extreme values, with B6 as the most challenging and F3 as the least, provide critical insights into task difficulty trends and their implications for AR interface design.

To validate these findings, vector data from a HoloLens experiment was integrated into Unity, enabling spatial mapping of user interactions. This mapping aligned with interaction quadrants, as shown in Figure 5.10. The analysis confirmed that specific quadrants, such as the upper frontal area, presented higher levels of task difficulty. These results are consistent with earlier findings, reinforcing the robustness of our methodology.

The findings from the WAD analysis and quadrant mapping strongly support the applicability of the optimized Fitts' Law model in real-world AR environments. The consistent alignment of experimental results with user feedback underscores the model's validity in capturing and predicting user interaction challenges in AR systems. This comprehensive evaluation assures the reliability of our approach and highlights the potential of Fitts' Law as a critical tool for understanding and optimizing AR user experiences.

5.4.3 Discussion

The findings from this study significantly advance the design of augmented reality (AR) interfaces and contribute to the broader field of human-computer interaction (HCI). Integrating head movement and spatial orientation into Fitts' Law provides a

TABLE 5.5: Weighted average task difficulty ratings for F, L, R, and B series targets based on user feedback.

F series		L series		R series		B series	
Target	WAD	Target	WAD	Target	WAD	Target	WAD
F1	2.38	L1	3.14	R1	2.86	B1	3.57
F2	2.85	L2	3.36	R2	3.43	B2	3.71
F3	2.50	L3	3.43	R3	3.57	B3	3.86
F4	4.35	L4	3.34	R4	3.71	B4	4.07
F5	4.21	L5	5.07	R5	5.00	B5	5.36
F6	4.35	L6	5.29	R6	5.36	B6	5.50

more accurate framework for predicting and optimizing interaction task difficulty in AR environments. A genetic optimization algorithm refined the model parameters, enabling adaptability across various AR scenarios. Moreover, using quaternion algebra to handle head rotations effectively addresses limitations in previous models of Fitts' Law.

The interaction space was structured around the user's head using frontal, left, right, and back anatomical planes subdivided into regions for detailed analysis (Figure 5.2). Results revealed varying levels of interaction task difficulty across these regions, with the upper frontal area being particularly challenging due to the complex gaze and head movements required for interactions above the transverse plane. This emphasizes the importance of ergonomic considerations, especially for objects positioned along the vertical axis, which pose additional challenges.

In contrast, the analysis showed that the lower frontal and lateral planes are more accessible, as they align more naturally with head positioning and gaze direction. However, task difficulty increased in peripheral areas due to the ergonomic constraints of head and neck movement. Incorporating head movement into the analysis introduced significant ergonomic nuances, with interaction task difficulty increasing in the initially faced quadrants (e.g., 1.51 for V1H1 with head influence, as shown in Figure 5.9).

To optimize the interaction model, a genetic algorithm was employed to adjust parameters such as Lambda (Λ), the strength of head movement (s), Sigma (σ), and Mu (μ). This process significantly enhanced the model's performance across spatial quadrants (Figure 5.3). The optimization began with baseline parameters ($\Lambda = 3$, $s = 6$, $\sigma = 8$, and $\mu = 4$) and yielded final optimized parameters ($\Lambda = 1.77$, $s = 0.07$, $\sigma = 0.60$, and $\mu = 0.02$). These adjustments achieved a minimized Mean Squared Error (MSE) of 0.98, reducing the average Level of Difficulty (LoD) by approximately 40%.

Quadrant-specific reductions in task difficulty were observed, with the first quadrant experiencing a 66.7% reduction, the fourth quadrant a 40.0% reduction, and the second and third quadrants reductions of approximately 33.3% and 20.0%, respectively (Figure 5.15). These results indicate a more consistent and ergonomic user experience across varying head distances.

Empirical validation demonstrated a strong correlation between predicted interaction times and actual user performance ($R^2 = 0.910$; Figure 5.12). Further analysis of the relationship between task complexity (LoD) and perceived task difficulty (WAD) confirmed the robustness of the model. This link between objective task metrics and user experience is illustrated in Figure 5.14.

Task Difficulty Distribution: Figure 5.16 presents the probability density of task difficulty levels across various user experiences within the AR environment after optimization. The normal distribution of task difficulty, with the mean (μ) indicating the average user-perceived challenge, confirms the model's improvements. The shaded areas represent 1-sigma (68.27%), 2-sigma (95.45%), and 3-sigma (99.73%) intervals, demonstrating the consistency of task difficulty levels post-optimization.

Our model demonstrated superior performance when compared to established methods, such as the raycasting technique evaluated by Mifsud et al. [196]. While raycasting achieved an average throughput of 2.63 bits/second and movement times of 1068 ms, our approach resulted in a 40% reduction in interaction task difficulty across quadrants. Including head movement and ergonomic considerations makes our model more suitable for complex AR environments, particularly in challenging areas like the upper frontal quadrant.

Additionally, recent work by Wagner et al. [194] highlights the advantages of gaze-hand alignment techniques, such as Gaze&Handray, which outperform raycasting for close-range tasks but struggle with distant targets due to parallax effects. Our model addresses these limitations by offering a versatile and efficient solution for AR interactions, effectively reducing task difficulty regardless of target distance.

The findings and refinements presented in this study provide a robust foundation for designing ergonomic AR interfaces. By integrating head movement and spatial orientation into predictive models, our approach ensures that interaction challenges are minimized across diverse scenarios. The improvements achieved through genetic optimization further enhance the usability and applicability of AR systems, paving the way for future innovations in user-centric interface design.

5.5 Conclusion

This chapter has presented a significant advancement in human-computer interaction (HCI) by adapting Fitts' Law for use in augmented reality (AR) environments. The adaptation process involved integrating head movement and spatial orientation into the traditional Fitts' Law model, aligning foundational HCI principles with the unique interaction dynamics of immersive technologies, including AR, virtual reality (VR), and mixed reality (MR). This adaptation deepens the understanding of user interaction in 3D spaces and extends the applicability of classic HCI models to next-generation interfaces.

The findings of this study demonstrate the potential of the adapted Fitts' Law model to reduce interaction task difficulty by approximately 40% across spatial quadrants. This reduction emphasizes the importance of designing intuitive, ergonomic AR systems by optimizing object placement and interface interaction. Such advancements hold relevance beyond AR, offering practical tools for developers to enhance usability in various immersive platforms. Moreover, the theoretical and experimental methodologies employed here establish a robust framework for evaluating and optimizing interactive systems across diverse technological environments.

In the broader context of this dissertation, the work contributes directly to the goal of enhancing user-centric design principles for immersive technologies. It underscores

the critical role of reimagining traditional HCI models to meet the demands of emerging interfaces, ensuring practical and intuitive interaction experiences across various platforms.

To further enrich the impact of these findings, future research should explore applying the adapted Fitts' Law model across a broader range of devices and immersive environments. While this study focused on Microsoft HoloLens 2, expanding to platforms such as Magic Leap, AR-enabled mobile devices, and VR headsets would provide a comparative analysis of interaction difficulties across various technologies. Such research would validate the generalizability and adaptability of the model, thereby strengthening its relevance to the field.

Additionally, integrating real-time cognitive load measurement techniques, such as EEG and eye-tracking, would provide deeper insights into how task difficulty affects mental workload in AR environments. This data could refine task difficulty metrics, further enhancing the accuracy and applicability of the model in designing interfaces for immersive technologies.

Therefore, this chapter's findings are pivotal in advancing the theoretical and practical understanding of interaction design in the rapidly evolving landscape of AR, VR, and MR technologies.

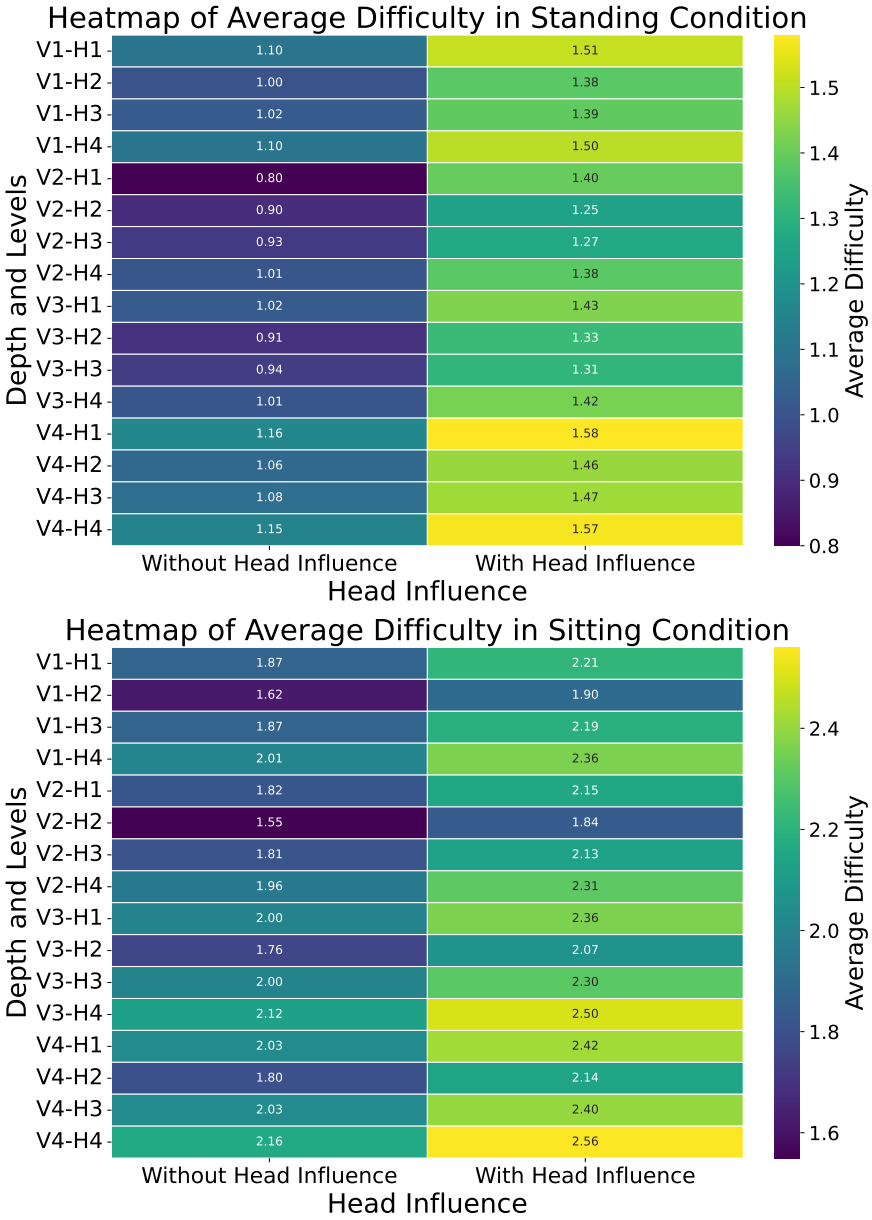


FIGURE 5.9: Heatmaps of AR Interaction Task Difficulty: These heatmaps visually contrast average task difficulty in standing and sitting positions, with (right) and without (left) head movement. Colour gradients indicate the difficulty level, with lighter shades denoting higher challenges. The visualization emphasizes how head movement increases task difficulty, providing insights for ergonomic AR design.

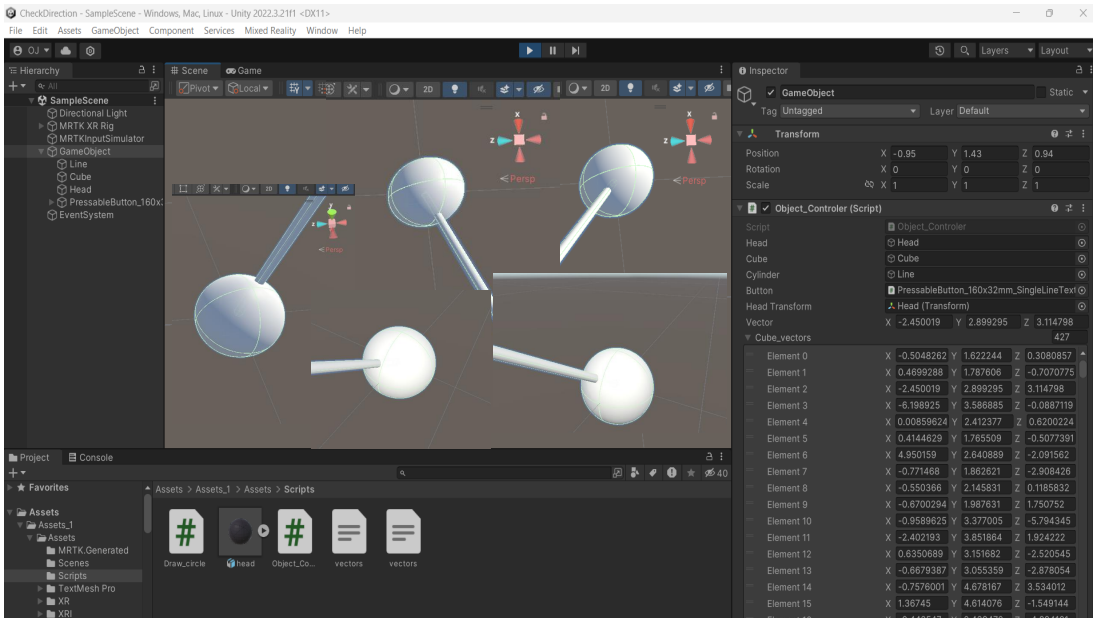


FIGURE 5.10: Vector Data from Unity showing interaction points and quadrants for empirical validation. Each vector represents a user's interaction within the AR environment, categorized into different spatial quadrants.

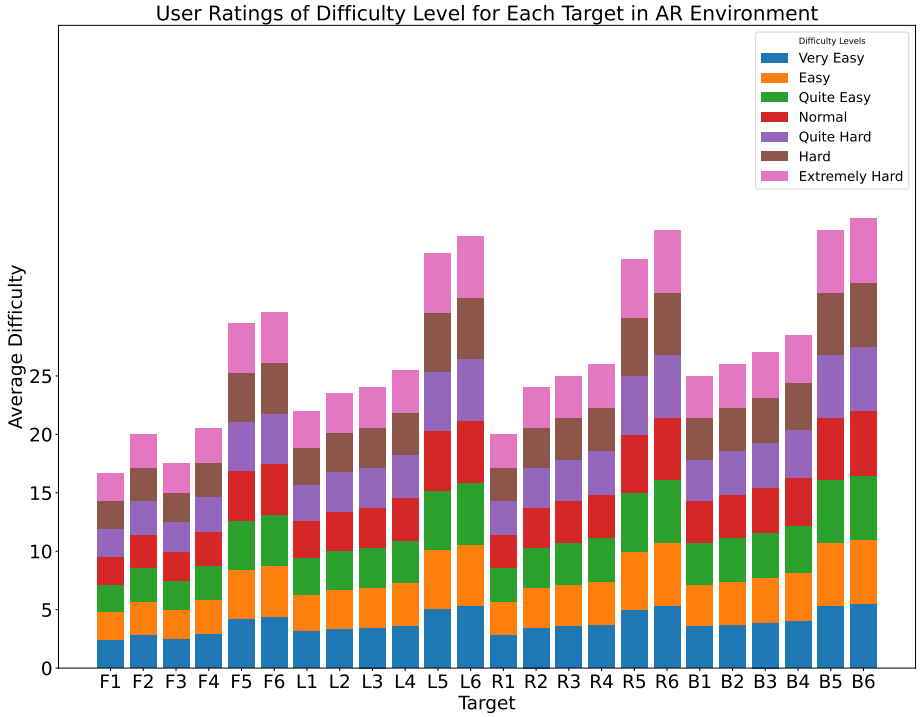


FIGURE 5.11: This chart presents the collective user feedback from a survey on interaction task difficulty within an augmented reality environment. It summarizes the relative frequency of perceived levels of difficulty, offering a consolidated view of user experiences across the spectrum of AR interactions.

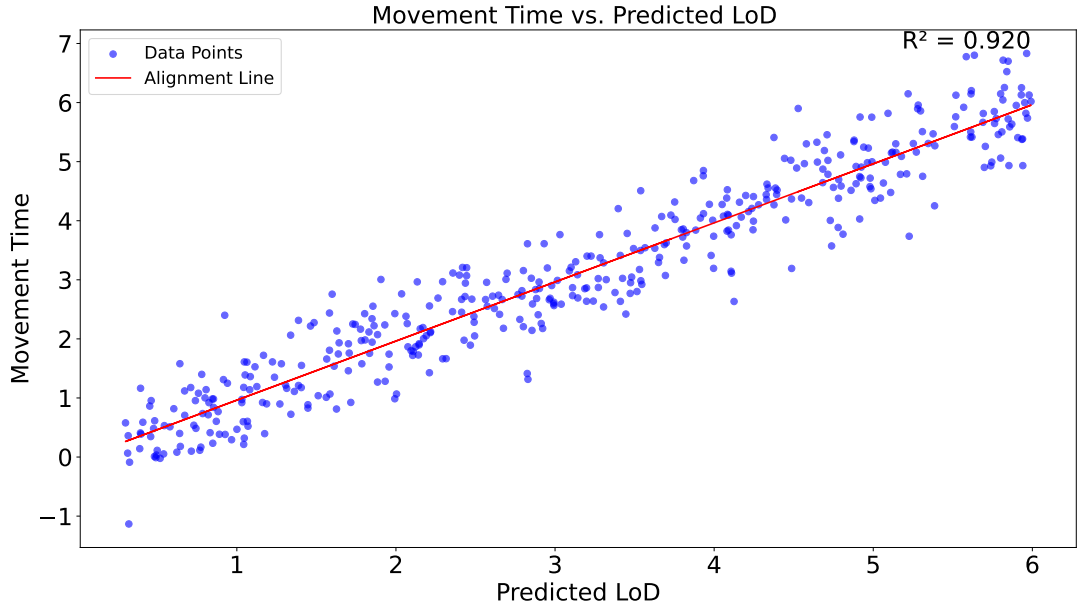


FIGURE 5.12: Scatter plot showing the correlation between Predicted LoD and actual user movement times, with an R^2 of 0.910, indicating the accuracy of the model.

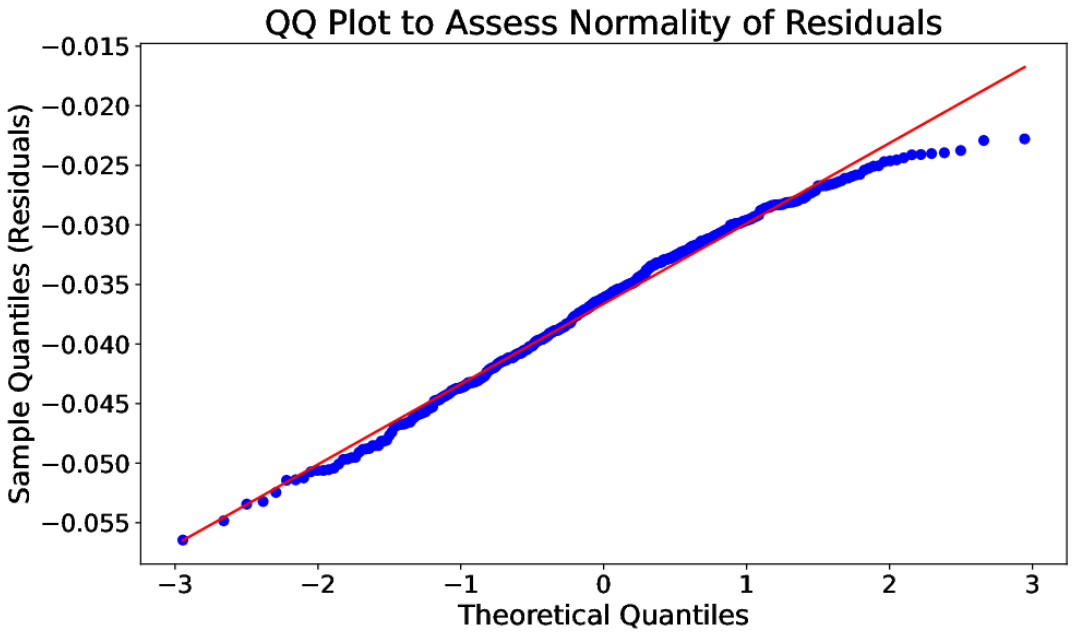


FIGURE 5.13: QQ Plot of residuals from the model fit, illustrating that the residuals closely follow a normal distribution, supporting the validity of the model's predictions

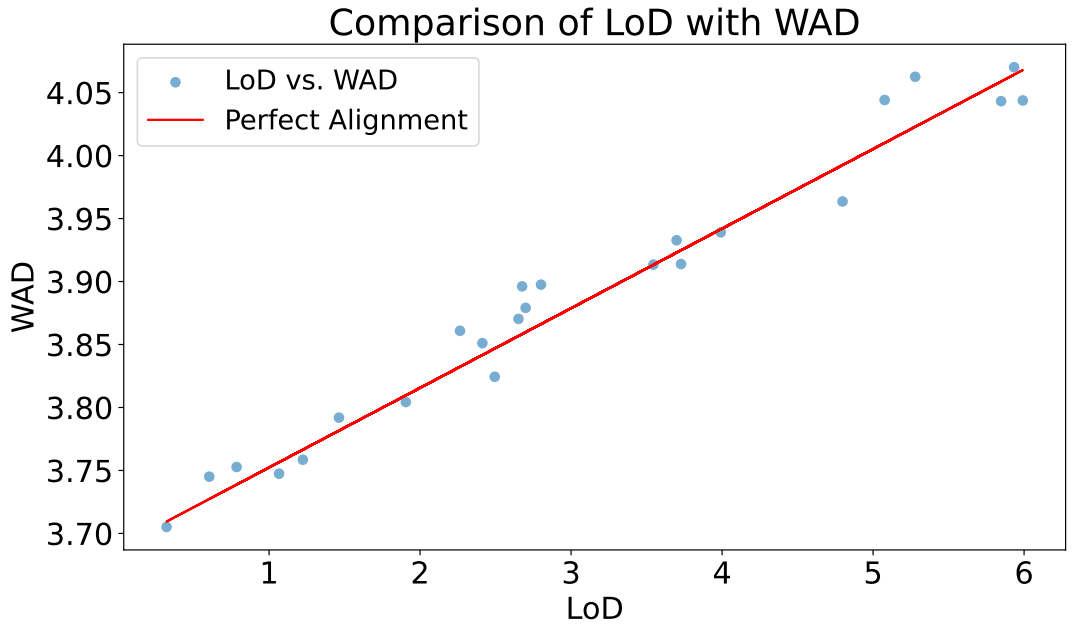


FIGURE 5.14: Correlation between Task Complexity (LoD) and User Perceived Task Difficulty (WAD) in an AR setting.

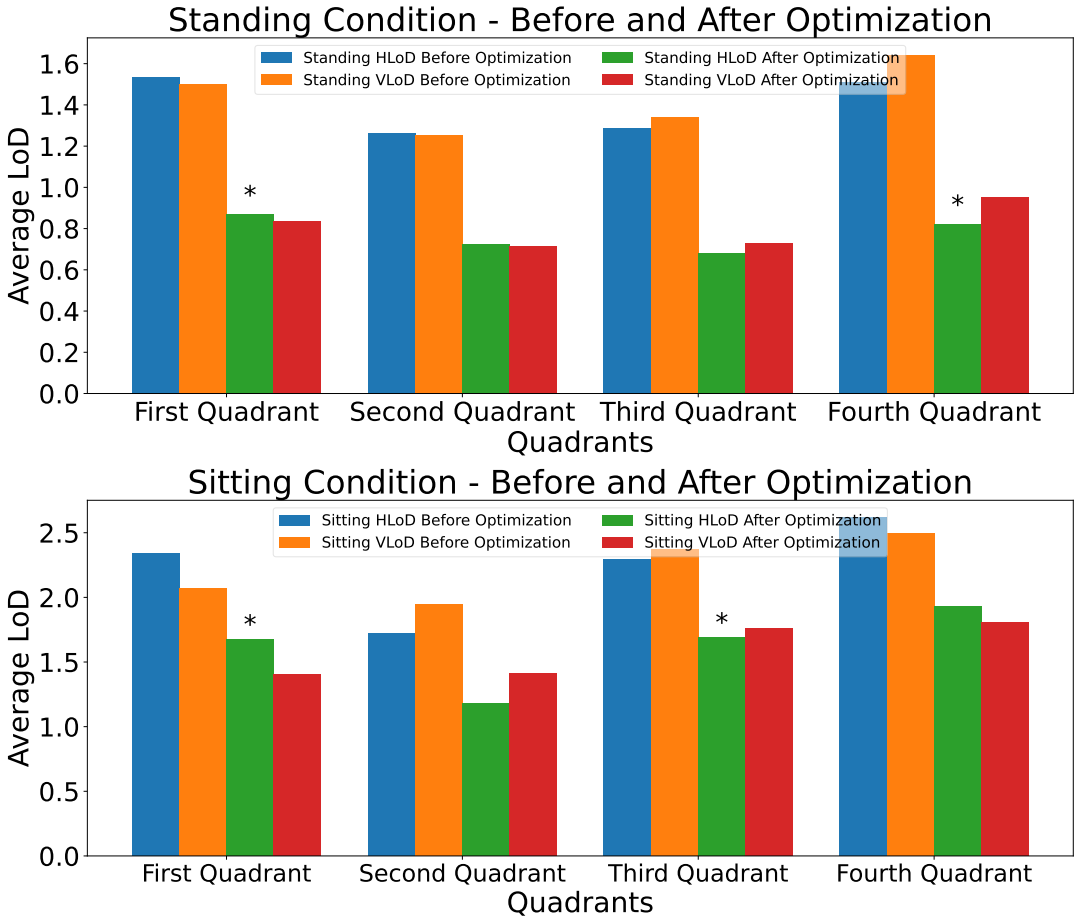


FIGURE 5.15: LoD Comparison by Quadrants, Pre- and Post-optimization

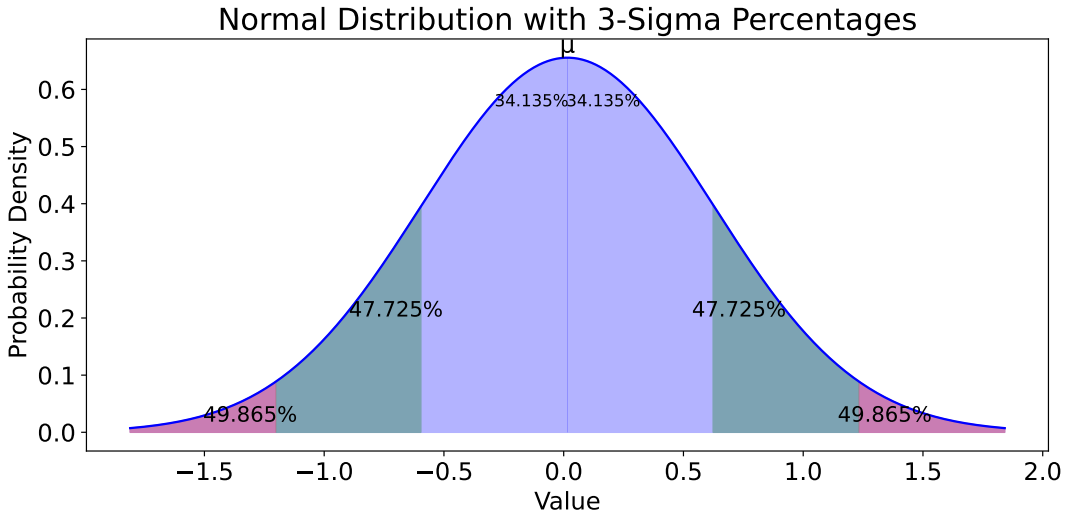


FIGURE 5.16: Probability density of task difficulty levels across AR interactions, showing the spread and likelihood of difficulty levels within one, two, and three standard deviation ranges.

6 Adaptive Kalman-Particle Filter Fusion for High-Fidelity Augmented Reality Head Tracking in Flight Simulation

6.1 Introduction

Training has become increasingly vital in augmented reality (AR) environments across industries where safety, precision, and efficiency are paramount. AR-based simulators offer immersive, risk-free environments that allow trainees to develop critical skills. Flight simulators enable pilots to practice essential skills in the aviation sector before transitioning to real aircraft. However, traditional flight simulators often lack the realism necessary to replicate the complexities of real-world scenarios, potentially limiting the effective transfer of knowledge and experience to trainees [151].

The advent of AR-based flight simulators represents a significant advancement in pilot training. These systems enhance realism by overlaying virtual elements onto physical environments, improving situational awareness and decision-making skills while reducing operational costs and risks [198]. Applications of AR-based simulation in aviation include aircraft maintenance [199], maintenance training [21], inspection [200], flight simulation [17], and emergency response [201]. For instance, AR overlays in maintenance can display diagrams and instructions, facilitating identification and repair [202], thereby improving efficiency and reducing errors. AR flight simulation allows pilots to practice various scenarios risk-free, enhancing situational awareness and decision-making skills [203]. Emergency response training benefits from AR by simulating fires, engine failures, hijackings, and landing manoeuvres [37]. Overall, AR simulators can accelerate training, addressing the increasing demand for skilled pilots within the aviation industry.

Beyond aviation, the adaptive Kalman-Particle Filter approach developed in this study holds potential for various augmented reality (AR) applications that require precise, real-time head tracking across diverse movement patterns and tasks. This filter fusion method, designed to balance computational efficiency and robustness dynamically, could be particularly valuable in high-stakes environments, such as medical simulations, where precise, controlled head movements are critical for tasks like surgical navigation [204]. In gaming, where users often make rapid, unpredictable head turns, the adaptive filter's ability to respond to varied motion profiles enhances its

[0]Publication note: This chapter is based on the peer-reviewed article [197]. http://dx.doi.org/10.1109/TC.2025.3566901.

applicability [205]. Similarly, automotive applications, especially in developing heads-up displays (HUDs) for driver assistance systems, could benefit from this method, which offers accurate, low-latency tracking even under variable driving conditions [206]. The flexibility of our filtering approach suggests that it could be adapted to meet the needs of these contexts by adjusting parameters in response to different sensor noise levels and motion dynamics, supporting a broader range of AR-based user experiences and tasks.

A critical technology underpinning the effectiveness of AR environments is head tracking. Head tracking, which involves real-time detection and monitoring of a user's head movements and orientation, is indispensable in applications such as Virtual Reality (VR), Augmented Reality (AR), gaming, human-computer interaction (HCI), and simulations. It enables intuitive control and immersive experiences, making accuracy and low latency essential for seamless interaction and user satisfaction, particularly in AR flight simulators for pilot training [102].

This dissertation addresses the challenges of latency and accuracy in head tracking within AR flight simulators. These challenges arise due to the complex requirements of real-time interaction in immersive environments. Latency, or the delay between physical movement and system response, can disrupt the user experience, potentially causing disorientation or motion sickness [207]. Inaccurate head tracking can lead to misaligned virtual elements, reducing the realism of simulations and hindering the development of critical spatial awareness and decision-making skills [208], [209]. Therefore, balancing low latency and high accuracy is essential for creating effective AR-based training environments.

6.1.1 Head Tracking in AR-Based Flight Simulation

Head tracking is critical for aligning virtual components with real-world actions, ensuring immersive and realistic AR experiences. It plays a central role in applications such as virtual reality (VR), augmented reality (AR), gaming, human-computer interaction (HCI), and simulations. In AR flight simulation, head tracking enables pilots to interact seamlessly with virtual instruments and observe their surroundings within a simulated cockpit, replicating the natural visual experience of piloting an aircraft [17].

Advanced AR platforms, such as Microsoft's HoloLens 2, exemplify the integration of technologies like Simultaneous Localization and Mapping (SLAM), which enable precise and low-latency head tracking. SLAM technology allows real-time mapping of the environment while tracking the user's position and orientation. By leveraging advanced sensors—including an inertial measurement unit (IMU), gyroscope, accelerometer, and depth cameras—the HoloLens 2 ensures accurate alignment between virtual and physical elements in AR simulations [37]. This integration is particularly critical in flight training scenarios, where precise head tracking allows pilots to practice various manoeuvres in a controlled and immersive environment.

Despite these advancements, achieving low latency and high accuracy simultaneously remains a significant challenge. Latency, or the delay between physical movement and system response, disrupts immersion and can cause motion sickness or disorientation, particularly during prolonged sessions [207], [210]. Variations in latency may also result in motion jumps, where the virtual environment fails to track user movements accurately. Similarly, inaccuracies in head tracking, often caused

by limitations in optical or inertial measurement systems, can lead to misalignment between virtual and physical elements, reducing the realism of the simulation and hindering skill development [208].

Existing head-tracking technologies, such as standalone Kalman and particle filters, address some challenges but face limitations in balancing latency and accuracy. Kalman filters offer fast processing but struggle with non-linear movements [211]. While capable of handling non-linearities, particle filters often incur higher computational demands, impacting real-time performance [212]. This trade-off highlights the need for advanced solutions that dynamically adapt to real-time factors and maintain optimal performance.

This dissertation introduces an adaptive Kalman-Particle Filter fusion approach to address these challenges. Unlike traditional methods that rely on static weighting or independent filter operations, this approach dynamically adjusts the contributions of the Kalman and particle filters based on real-time factors such as sensor noise levels and head movement characteristics. By incorporating latency compensation techniques within the Kalman filter framework and optimizing filter parameters for the AR flight simulator environment, the proposed method aims to minimize latency, reduce jitter, and improve tracking accuracy. These innovations are essential for ensuring a seamless and responsive AR experience in flight simulation while addressing the specific needs of pilot training environments.

6.1.2 Review of related works

Despite AR systems' potential, accurate head tracking remains a critical challenge, especially in scenarios requiring real-time responsiveness, such as flight training [54]. Precise head tracking enables the alignment of virtual and physical elements, which is essential for creating an immersive experience. This technology relies on real-time head movement and orientation monitoring, allowing for intuitive control and natural interaction with virtual environments. In AR flight simulations, tracking sensors such as optical and inertial measurement units (IMUs) provide essential data for calculating the pilot's head position and orientation. However, optical sensors can suffer from occlusion, and inertial sensors are prone to drift, creating inaccuracies in tracking [213]. These limitations can disrupt the alignment between real and virtual elements, leading to latency, jitter, and tracking errors, which degrade the pilot's experience and introduce risks like motion sickness, often referred to as cybersickness [207]. Latency—the delay between head movement and the system's visual response—can disorient users and induce discomfort, particularly during extended training sessions.

Several researchers have applied Kalman and particle filters to head tracking in AR/VR systems to address these challenges. Ababsa et al. [214] demonstrated particle filters' capacity to handle non-linear and non-Gaussian noise in head tracking applications, making them ideal for reducing latency and improving dynamic registration within augmented reality (AR) head-mounted displays (HMDs). By employing a sequential Monte Carlo method, particle filters provide a robust alternative to the Extended Kalman Filter (EKF), effectively managing complex head motion patterns. Ababsa and Malle [215] also showed that particle filters achieved lower RMSE values in both position and orientation, outperforming EKF regarding tracking accuracy, especially under conditions with high noise and non-linear motion.

Moreover, Gül et al. [61] demonstrated the benefits of a Kalman Filter-based approach for cloud-based mixed reality, where latency is compounded by remote rendering demands. Their study found that the Kalman filter effectively compensates for motion-to-photon latency in cloud-rendered systems, improving head orientation prediction accuracy and thus minimizing visual lag for users engaged in real-time AR applications.

Bouaynaya et al. [216] introduced a motion-based particle filter specifically designed for complex head-tracking environments. Their approach enhanced traditional particle filtering by employing Adaptive Block Matching (ABM) as a motion estimation technique, allowing the system to adapt efficiently to abrupt and erratic head movements—such as quick rotations or jumps—by improving the distribution of particles according to the observed motion. This adaptation allowed the particle filter to handle sudden movements that are challenging for conventional tracking models, such as the CONDENSATION filter, thereby increasing tracking reliability in dynamic AR/VR settings.

Beyond filter-based tracking, SLAM-based methods leverage visual-inertial odometry to map the environment and localize the user's head concurrently. This enhances robustness in cluttered or occluded indoor settings and dynamic outdoor environments, as demonstrated by Microsoft HoloLens 2 [16]. Deep learning methods, such as CNNs and RNNs, achieve high accuracy in head pose estimation [217]. However, they often require substantial computational resources and extensive training data, limiting their suitability for lightweight AR hardware. RTK-GPS offers centimetre-level positional accuracy in outdoor settings, but it is ineffective indoors due to its reliance on satellite signals and lack of orientation data.

Recent advancements have focused on improving AR registration accuracy under strict attitude rotation constraints. These include optimization-based methods and hybrid inertial-visual fusion techniques that ensure stable pose estimation during rapid head turns or off-axis rotations [218]. Cross-platform AR tracking frameworks now support consistent head tracking across a wide range of hardware and operating systems, including mobile and wearable AR devices [219]. Additionally, Transformer-based models have shown strong potential in head pose estimation by capturing long-range dependencies and temporal features from sensor or video input streams [220]. These approaches outperform traditional CNNs in sequence modeling and offer a promising direction for robust head tracking in real-time AR applications. In contrast, the proposed Adaptive Kalman-Particle Filter (AKPF) balances computational efficiency, real-time adaptability, and robustness, making it well-suited for resource-constrained AR platforms in immersive training scenarios.

Building upon this work, this research proposes an adaptive Kalman-particle filter fusion strategy tailored explicitly for head tracking in AR flight simulation environments using Microsoft HoloLens 2. By dynamically adjusting filter parameters and integrating optimizations suited to the challenges of AR flight simulation, this approach achieves low latency and high tracking accuracy, ensuring reliable performance under diverse and demanding conditions.

6.1.3 The main contributions of this chapter

In our study, we use Microsoft HoloLens 2, which incorporates SLAM technology. By utilizing IMUs, gyroscopes, and depth cameras, SLAM enhances head tracking by mapping the environment in real-time and providing precise positioning data [16]. However, achieving low-latency and high-accuracy head tracking remains a significant technical challenge.

These studies underscore the potential of both Kalman and particle filters for addressing the unique head-tracking challenges posed by AR/VR applications. However, while Kalman filters are computationally efficient for linear tracking, they struggle with non-linear motion [211]; conversely, particle filters handle non-linear dynamics well but require significant computational resources, which may increase latency [212]. We propose an adaptive fusion of Kalman and particle filters to bridge these limitations, a novel approach to achieve low-latency, high-accuracy tracking. Our adaptive fusion method dynamically balances the strengths of Kalman filters' computational efficiency with the particle filters' robustness in non-linear tracking, based on real-time factors such as sensor noise and head movement patterns.

The key contributions of this study are as follows:

- **Adaptive Kalman-Particle Filter Fusion**: We introduce a dynamic fusion of the Kalman and Particle filters that adjusts contributions based on real-time conditions, such as sensor noise and head movement patterns. This approach addresses the limitations of existing standalone filters that either struggle with non-linear motion or incur high computational costs and latency.

- **Low-Latency Head Tracking**: Our fusion approach incorporates latency compensation techniques that significantly reduce system lag, ensuring real-time responsiveness. This is essential for preventing motion sickness and enhancing user experience in pilot training.

- **Application-Specific Optimization**: We introduce a novel, domain-specific optimization of filtering parameters tailored to AR-based flight simulators' unique head movement dynamics and sensor noise characteristics. Unlike conventional implementations that use static parameters, our approach dynamically tunes key variables (e.g., particle count, noise covariance) in response to real-time motion, enabling both tracking robustness and computational efficiency on resource-constrained devices like the HoloLens 2.

- **Comprehensive Evaluation of Filtering Techniques**: We evaluate various filtering techniques, including Extended Kalman Filter, Sequential Importance Resampling, Unscented Kalman Filter, Unscented Particle Filter, Ensemble Kalman Particle Filter, and Auxiliary Particle Filter. This evaluation provides valuable insights into their performance under real-time AR conditions.

By reducing jitter, latency, and tracking errors, our fusion approach enables pilots to engage in longer training sessions with minimized risk of motion sickness, ultimately contributing to more effective skill development in aviation training. Furthermore, the findings of this study hold potential for improving AR-based flight simulations and broader applications in other AR/VR environments, including healthcare, gaming, and HCI, where precise head tracking and user comfort are critical.

6.2 Materials And Methods

This chapter outlines the methodology used to develop and evaluate the Adaptive Kalman-Particle Filter (AKPF) fusion approach for head tracking in AR-based flight simulators. This work directly addresses the dissertation goal of enhancing real-time head tracking accuracy and latency in augmented reality (AR) systems for flight simulators. By integrating the AKPF with SLAM capabilities of the Microsoft HoloLens 2, this study advances the fidelity and effectiveness of AR in aviation training, as outlined in Chapter 1.

6.2.1 Implementation in AR Flight Simulation

The AKPF was integrated into a custom-built flight simulation platform, leveraging Microsoft HoloLens 2 as the primary AR interface. This environment facilitates real-time performance evaluation, measuring latency, precision, and stability under dynamic conditions. Algorithm 1 and Figure 6.1 illustrate the iterative fusion process, highlighting the transition between Kalman and particle filters during dynamic head tracking. To foster reproducibility, the complete implementation and performance benchmarks are available on GitHub.

AR-based flight simulation presents unique technical demands that set it apart from other AR domains. These include constrained head movement within cockpit environments, rapid gaze shifts between visual instruments and external overlays, and an acute sensitivity to latency and jitter. Even small tracking errors can disrupt a trainee's perception of the aircraft's state or external environment, thereby reducing training effectiveness. The AKPF algorithm is explicitly designed to meet these challenges through dynamic tuning of Kalman and particle filter contributions based on real-time motion intensity and sensor feedback, ensuring precise and responsive tracking tailored to the needs of flight simulation.

6.2.2 Adaptive Kalman-Particle Filter Fusion Methodology

The Adaptive Kalman-Particle Filter (AKPF) fusion methodology combines the complementary strengths of Kalman and Particle filters, leveraging their distinct advantages in linear and nonlinear motion tracking. Kalman filters are highly efficient for error correction in linear systems, providing computational efficiency and effective noise reduction. In contrast, Particle filters excel at handling nonlinear and non-Gaussian dynamics but are computationally intensive. The AKPF dynamically balances these strengths to enable robust and accurate head tracking, even in challenging conditions like AR-based flight simulators.

To evaluate the feasibility of the AKPF for real-time applications, its computational performance was analyzed on the Microsoft HoloLens platform. The algorithm achieves real-time tracking with minor optimizations, thanks to the HoloLens's processing capabilities. While adaptive particle resampling—an inherent component of Particle filters—requires significant computational resources, the efficiency of the Kalman filter in error correction offsets this load. This balance ensures smooth operation on the HoloLens hardware and maintains tracking accuracy across varied conditions. The AKPF's adaptive design dynamically adjusts to noise levels and

abrupt movements, ensuring robust real-time performance critical for maintaining immersive AR experiences.

A key feature of the AKPF is its flexibility in adjusting critical parameters—such as particle count, process noise (Q), and measurement noise covariance (R)—to match the demands of different AR applications. For example, in low-noise environments, the particle count can be reduced to conserve computational resources, while in high-motion or noisy scenarios, these parameters are fine-tuned to maintain performance. This adaptability enables the AKPF to accommodate diverse head movement patterns and sensor noise conditions, making it highly suitable for flight simulation and other applications such as medical AR systems requiring precise and stable tracking or gaming contexts involving unpredictable head movements.

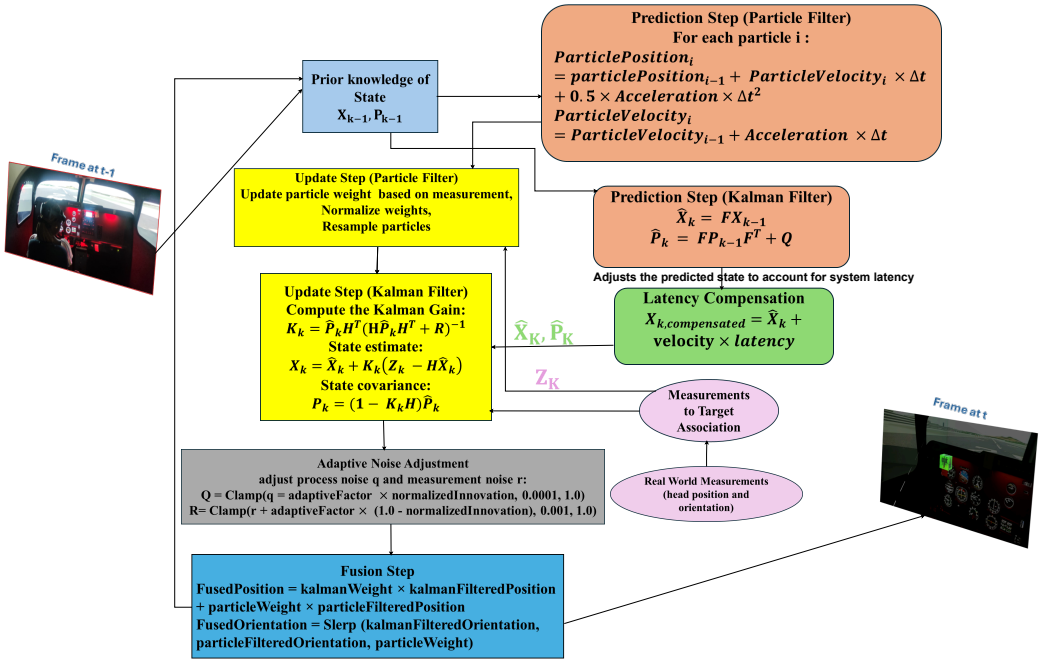


FIGURE 6.1: Schematic of Kalman and Particle Filter Fusion with Latency Compensation in a Cockpit.

The following subsections detail the methodologies employed for each filter type, which contribute to the AKPF framework.

6.2.3 Advanced Extended Kalman Filtering for Precise Head Tracking

The Extended Kalman Filter (EKF) is an advanced state estimation technique essential for non-linear systems such as head tracking in AR-based flight simulators. This adaptation builds upon the traditional EKF framework, optimizing it for the high dynamism and precision required in augmented reality applications. By employing a refined process model and customized measurement updates, this tailored EKF

Algorithm 2 Adaptive Kalman Particle Filter Fusion Framework for State Estimation

Inputs: Number of particles N, process noise covariance Q, measurement noise covariance R, state transition matrix F, measurement matrix H, initial state $\mathbf{x}_0$, initial covariance P_0, control input $\mathbf{u}$, measurements $\mathbf{z}$
Outputs: Fused state estimate $\hat{\mathbf{x}}$
procedure INITIALIZE
    Initialize Kalman filter parameters $Q, R, F, H, \mathbf{x}_0, P_0$
    Initialize particle filter with N particles with initial states $\mathbf{x}_i$ and weights $w_i = \frac{1}{N}$
    Application-Specific Optimization:
    Tailor Q and R based on the AR flight simulator's head movement patterns and sensor noise profiles
    Implement focused resampling strategy to identify regions of interest based on recent head movement patterns
end procedure
procedure PREDICT
    Kalman Predict:
    $\mathbf{x}_{k|k-1} \leftarrow F\mathbf{x}_{k-1|k-1} + \mathbf{u}_{k-1}$
    $P_{k|k-1} \leftarrow FP_{k-1|k-1}F^T + Q$
    Particle Predict:
    for each particle i **do**
        $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i\Delta t + \frac{1}{2}\mathbf{a}\Delta t^2$
        $\mathbf{v}_i \leftarrow \mathbf{v}_i + \mathbf{a}\Delta t$
    end for
end procedure
procedure UPDATE
    Kalman Update:
    $K_k \leftarrow P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
    $\mathbf{x}_{k|k} \leftarrow \mathbf{x}_{k|k-1} + K_k(\mathbf{z}_k - H\mathbf{x}_{k|k-1})$
    $P_{k|k} \leftarrow (I - K_kH)P_{k|k-1}$
    Particle Update:
    for each particle i **do**
        $w_i \leftarrow w_i \cdot p(\mathbf{z}_k|\mathbf{x}_i)$
    end for
    Normalize particle weights w_i
end procedure
procedure FUSION
    $\hat{\mathbf{x}} \leftarrow \alpha\mathbf{x}_{k|k} + (1-\alpha)\sum_{i=1}^{N} w_i\mathbf{x}_i$
    // The value of α is determined based on the confidence in the Kalman filter and particle filter estimates
end procedure
procedure LOGGING AND ADJUSTMENT
    Log tracking data: position, orientation, and velocities
    Adjust process and measurement noise parameters Q and R based on innovation
end procedure

approach effectively manages the stochastic nature of rapid head movements, crucial for minimizing latency and enhancing the immersive experience in AR settings.

Adapted Process Model for Head Tracking: In the specific context of head tracking for AR flight simulation, the state vector $\mathbf{x}_k$ is comprehensively defined to capture both the linear and angular dynamics of head movements, crucial for maintaining a synchronized virtual environment:

$$\mathbf{x}_k = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} & \phi & \theta & \psi \end{bmatrix}^\top \tag{6.1}$$

Where:

- x, y, z: Represent the 3D position of the user's head.
- $\dot{x}, \dot{y}, \dot{z}$: Linear velocities in the respective axes.
- ϕ, θ, ψ: Roll, pitch, and yaw angles representing the head orientation.

The adapted state transition model incorporates advanced kinematic equations, accounting for the complex interactions between translational and rotational motions:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{u}_k + \mathbf{w}_k \tag{6.2}$$

Here, $\mathbf{F}_k$ is the state transition matrix, $\mathbf{u}_k$ the control input reflecting user-initiated movements, and $\mathbf{w}_k$ the process noise vector, modeled as Gaussian to adequately represent the uncertainty in head motion dynamics.

Enhanced Measurement Model: The measurement update step is critical for refining the state estimates using actual sensor data ($\mathbf{z}_k$), which in the context of AR includes data from optical and inertial sensors:

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^- \tag{6.3}$$

where $\mathbf{y}_k$ is the measurement residual, and $\mathbf{H}_k$ is the measurement matrix, meticulously designed to link the state variables to the measurements obtained from AR hardware.

The Kalman gain $\mathbf{K}_k$, crucial for updating the estimates, is computed as:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^\top + \mathbf{R})^{-1} \tag{6.4}$$

Here, $\mathbf{R}$ represents the measurement noise covariance matrix, which is fine-tuned to reflect the unique characteristics of sensors employed in AR systems for flight simulation.

Optimization with Jacobian Linearization: Given the non-linear nature of head movement dynamics, the EKF employs Jacobian matrices to linearize the non-linearities around the current estimate:

$$\mathbf{F}_{k-1} = \frac{\partial f}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}} \tag{6.5}$$

$$\mathbf{H}_k = \frac{\partial h}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_k} \tag{6.6}$$

For a rigorous theoretical foundation, this methodology builds upon the standard EKF formulas tailored for AR applications, with specific adaptations for head tracking

as described in [221] and [222]. This approach ensures a high degree of accuracy and responsiveness, addressing the critical requirements of AR-based flight simulation environments.

6.2.4 Enhanced Sequential Importance Resampling (SIR) for AR Head Tracking

The Sequential Importance Resampling (SIR) particle filter, a robust Monte Carlo-based algorithm, has been refined explicitly for state estimation in the nonlinear and non-Gaussian environments characteristic of AR-based flight simulators. This dissertation presents an advanced adaptation of the SIR particle filter designed to handle the high sensor noise and complex, dynamic head movements typical in augmented reality, where real-time, accurate head tracking is essential.

Adaptive Process Model: The state transition model has been intricately tailored to include comprehensive kinematics and quaternion-based orientation updates, ensuring high fidelity in head movement tracking:

$$\mathbf{x}_k^i = \begin{bmatrix} \mathbf{r}_k^i \\ \mathbf{v}_k^i \\ \mathbf{q}_k^i \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{k-1}^i + \mathbf{v}_{k-1}^i \Delta t \\ \mathbf{v}_{k-1}^i + \mathbf{a}_{k-1} \Delta t \\ \mathbf{q}_{k-1}^i \otimes \boldsymbol{\Delta q} \end{bmatrix} + \mathbf{w}_{k-1}^i \tag{6.7}$$

Where:

- $\mathbf{r}_k^i$: Position (x, y, z).
- $\mathbf{v}_k^i$: Velocity $(\dot{x}, \dot{y}, \dot{z})$.
- $\mathbf{q}_k^i$: Quaternion for orientation, enhancing rotational accuracy.
- Δt: Time step optimized for minimal latency in dynamic simulations.
- $\boldsymbol{\Delta q}$: Quaternion increment, encapsulating detailed rotational dynamics.
- $\mathbf{w}_{k-1}^i$: Process noise, modeled with a refined covariance matrix $\mathbf{Q}$ to accurately reflect the sensor noise characteristics in AR settings [223].

Enhanced Quaternion Update Mechanism: Quaternion updates are critical for maintaining orientation accuracy, implemented using:

$$\mathbf{q}_k^i = \mathbf{q}_{k-1}^i \otimes \boldsymbol{\Delta q} \tag{6.8}$$

The quaternion multiplication $\otimes$ ensures seamless rotational transitions, vital for a jitter-free user experience in augmented environments.

Advanced Measurement Update: The measurement model incorporates high-definition RGB-D camera data, crucial for integrating depth and positional information with unprecedented precision:

$$p(\mathbf{z}_k | \mathbf{x}_k^i) = \exp\left(-\frac{1}{2} (\mathbf{z}_k - \mathbf{H}\mathbf{x}_k^i)^\top \mathbf{R}^{-1} (\mathbf{z}_k - \mathbf{H}\mathbf{x}_k^i) \right) \tag{6.9}$$

Where:

- **H**: Customized measurement matrix, meticulously designed to map state variables to sensor data, enhancing the fidelity of sensor integration.
- $\mathbf{z}_k$: Advanced sensor measurement vector, capturing both positional and depth data.
- **R**: Measurement noise covariance matrix, specially calibrated for the specific noise characteristics of RGB-D sensors in AR environments.

GPU-Accelerated Resampling in SIR Particle Filter

Resampling is a pivotal phase in the Sequential Importance Resampling (SIR) particle filter where particles are selectively propagated based on their normalized weights. This step is crucial for mitigating the degeneracy problem, which occurs when only a few particles end up dominating the weight distribution. To achieve real-time performance and maintain a representative set of particles, resampling is executed on GPUs, leveraging parallel processing capabilities efficiently.

Systematic Sampling and Weight Normalization: To begin the resampling process, each particle's weight is first normalized to ensure that the sum of all weights equals one:

$$\hat{w}_k^i = \frac{w_k^i}{\sum_{j=1}^{N} w_k^j} \tag{6.10}$$

Subsequently, cumulative weights are calculated for each particle, forming a distribution over which particles will be resampled:

$$C_k^i = \sum_{j=1}^{i} \hat{w}_k^j \tag{6.11}$$

The systematic sampling mechanism then selects new particles based on these cumulative weights and a uniformly drawn random number u, ensuring that particles with higher weights are more likely to be chosen. The selection criterion is described by:

$$\mathbf{x}_k^i = \mathbf{x}_k^j, \quad \text{where } C_k^{j-1} < u \leq C_k^j \tag{6.12}$$

For instance, in a simplified scenario with five particles where the normalized weights after the weight update step are [0.1, 0.1, 0.1, 0.1, 0.6], the cumulative sums of these weights would be [0.1, 0.2, 0.3, 0.4, 1.0]. If a random number $u = 0.35$ is drawn, the fourth particle is selected for propagation into the next generation because $0.3 < 0.35 \leq 0.4$. This systematic approach is repeated for each particle in the set, effectively preventing the concentration of weights in a few particles and maintaining a diverse set of hypotheses.

By implementing this GPU-accelerated systematic resampling, the SIR Particle Filter optimizes the computational efficiency and accuracy of the state estimation in real-time applications, such as head tracking in AR systems. This method not only preserves the diversity of the particle set but also ensures that the computational

resources are focused on the most probable states, enhancing the performance and reliability of the filter in dynamic environments.

Optimized Noise Tuning: The noise parameters ($\mathbf{Q}, \mathbf{R}$) are meticulously tuned based on extensive empirical data from AR head-tracking experiments, ensuring the filter's sensitivity and responsiveness are ideally suited for the dynamic and unpredictable nature of AR applications:

$$\mathbf{Q} = \sigma_q^2 \mathbf{I}, \quad \mathbf{R} = \sigma_r^2 \mathbf{I} \tag{6.13}$$

Where σ_q^2 and σ_r^2 represent empirically derived variances that optimize tracking performance under typical operational conditions.

This advanced implementation of the SIR particle filter addresses the unique challenges of AR-based head tracking by combining detailed kinematic modelling, sophisticated noise management, and high-performance computational techniques. The result is a highly accurate, low-latency, and jitter-free head-tracking system essential for immersive AR experiences in flight simulators. For a deeper understanding of the SIR particle filter and its foundational principles, refer to seminal works such as Doucet et al. [224] and Arulampalam et al. [223]. These references provide a comprehensive insight into the theoretical underpinnings and real-world implementations of particle filtering techniques, affirming the methodologies adapted in this dissertation for enhancing head tracking accuracy in AR systems.

6.2.5 Advanced Unscented Kalman Filter (UKF) for Head Tracking in AR Flight Simulators

The Unscented Kalman Filter (UKF) is a superior recursive algorithm designed to estimate the state of dynamic systems from noisy observations accurately. Leveraging a deterministic sampling technique known as the sigma points method, the UKF surpasses traditional Kalman filters by avoiding the need to linearize highly nonlinear dynamics, such as those encountered in AR-based flight simulators for head tracking. This section outlines the UKF modifications adapted for improved quaternion-based orientation updates and advanced sensor integration.

State and Observation Models The UKF utilizes two critical models: the state transition and the observation models. The state transition model describes the system's evolution over time, and the observation model links the system's state to actual measurements obtained from sensors. These models are pivotal for predicting the system's behaviour and updating the estimates based on new sensor inputs. The augmented state vector $\mathbf{x}_k$ now includes quaternion components for precise orientation tracking:

$$\mathbf{x}_k = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} & q_w & q_x & q_y & q_z \end{bmatrix}^\top \tag{6.14}$$

Process Model The adapted process model now includes detailed quaternion updates to account for rotational dynamics accurately, which are essential for AR applications:

$$f(\mathbf{x}_{k-1}) = \begin{bmatrix} \mathbf{r}_{k-1} + \mathbf{v}_{k-1}\Delta t \\ \mathbf{v}_{k-1} + \mathbf{a}_{k-1}\Delta t \\ \mathbf{q}_{k-1} \otimes \mathbf{\Delta q} \end{bmatrix} \tag{6.15}$$

Measurement Model To enhance measurement accuracy, the model now leverages advanced RGB-D sensor data, which provides both positional and depth information:

$$h(\mathbf{x}_k) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{6.16}$$

Modified UKF Equations

Sigma Points Generation: Sigma points are selected to capture the posterior distribution's mean and covariance effectively:

$$\chi_0 = \mathbf{x} \tag{6.17}$$

$$\chi_i = \mathbf{x} + \sqrt{(L+\lambda)\mathbf{P}}_i \quad \text{for } i = 1, \ldots, L \tag{6.18}$$

$$\chi_{i+L} = \mathbf{x} - \sqrt{(L+\lambda)\mathbf{P}}_i \quad \text{for } i = 1, \ldots, L \tag{6.19}$$

Sigma Points Propagation Through Process Model: Each sigma point is propagated through the nonlinear state transition function to predict the following state:

$$\chi_i' = f(\chi_i) \quad \text{for } i = 0, \ldots, 2L \tag{6.20}$$

Predicted State and Covariance: The predicted state and its covariance are calculated by recombining the propagated sigma points:

$$\hat{\mathbf{x}}_k = \sum_{i=0}^{2L} W_i^{(m)} \chi_i' \tag{6.21}$$

$$\hat{\mathbf{P}}_k = \sum_{i=0}^{2L} W_i^{(c)} (\chi_i' - \hat{\mathbf{x}}_k)(\chi_i' - \hat{\mathbf{x}}_k)^\top + \mathbf{Q} \tag{6.22}$$

Sigma Points Propagation Through Measurement Model: Sigma points are then used to predict the following measurement:

$$\zeta_i = h(\chi_i') \quad \text{for } i = 0, \ldots, 2L \tag{6.23}$$

Predicted Measurement and Innovation Covariance: The predicted measurements are used to compute the innovation covariance:

$$\hat{\mathbf{z}}_k = \sum_{i=0}^{2L} W_i^{(m)} \zeta_i \tag{6.24}$$

$$S_k = \sum_{i=0}^{2L} W_i^{(c)} (\zeta_i - \hat{\mathbf{z}}_k)(\zeta_i - \hat{\mathbf{z}}_k)^\top + \mathbf{R} \tag{6.25}$$

Kalman Gain and State Update: The Kalman Gain is derived to minimize the error between the predicted and actual measurements, updating the state estimate:

$$K_k = T_k S_k^{-1} \tag{6.26}$$
$$\mathbf{x}_k = \hat{\mathbf{x}}_k + K_k(\mathbf{z}_k - \hat{\mathbf{z}}_k) \tag{6.27}$$

Covariance Update: The covariance of the estimate is updated to reflect the latest knowledge:

$$\mathbf{P}_k = \hat{\mathbf{P}}_k - K_k S_k K_k^\top \tag{6.28}$$

By propagating sigma points through the nonlinear process and measurement functions, the UKF provides robust state estimates. This capability is critical in head-tracking applications where the quaternion-based process model allows the filter to adjust effectively for rotational dynamics, and the integration with RGB-D sensors ensures precise positional tracking. For a comprehensive understanding of the UKF's mathematical foundation and process model, see seminal works by Eric and Rudolph [225], Simon [226], and further developments [227] [222].

6.2.6 Advanced Ensemble Kalman Filter (EnKF) for Precise Head Tracking in AR Flight Simulators

The Ensemble Kalman Filter (EnKF) represents a sophisticated adaptation of the classic Kalman filter, employing a Monte Carlo approach to manage nonlinear and non-Gaussian state estimations effectively. This research utilizes the EnKF to refine the head position and orientation estimation within a flight simulator, integrating quaternion-based updates and RGB-D sensory data to address the complex dynamics typically encountered in augmented reality (AR) applications. The methodologies employed here draw upon the foundational principles laid out in [228] and [229], with significant enhancements to cater to the specific demands of real-time AR tracking.

State Representation In this study, the state vector is comprehensively designed to encapsulate all necessary motion descriptors:

$$\mathbf{x}_k = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} & q_w & q_x & q_y & q_z \end{bmatrix}^\top \tag{6.29}$$

Where x, y, z represent the 3D position coordinates, $\dot{x}, \dot{y}, \dot{z}$ denote the linear velocities, and q_w, q_x, q_y, q_z are the quaternion components crucial for representing the orientation of the user's head.

Process Model The process model is intricately formulated to incorporate both translational and rotational dynamics:

$$\mathbf{x}_k^{(i)} = \begin{bmatrix} \mathbf{r}_k^{(i)} + \mathbf{v}_k^{(i)} \Delta t \\ \mathbf{v}_k^{(i)} + \mathbf{a}_k^{(i)} \Delta t \\ \mathbf{q}_k^{(i)} \otimes \Delta \mathbf{q}^{(i)} \end{bmatrix} + \mathbf{w}_k^{(i)} \tag{6.30}$$

This model takes into account the instantaneous position $\mathbf{r}_k^{(i)}$, velocity $\mathbf{v}_k^{(i)}$, and the quaternion updates $\Delta \mathbf{q}^{(i)}$, ensuring dynamic responsiveness to user movements with the addition of process noise $\mathbf{w}_k^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ to model uncertainties.

Measurement Model The measurement model utilizes RGB-D sensors to obtain accurate spatial data:

$$\mathbf{z}_k = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{v}_k \tag{6.31}$$

Here, $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ represents the measurement noise, accounting for the potential discrepancies introduced by sensor inaccuracies.

EnKF Algorithm: Comprehensive Approach

Prediction Step: Each ensemble member is propagated through the process model to predict the next state, incorporating control inputs and modelling uncertainties:

$$\mathbf{x}_k^{(i)} = \mathbf{f}(\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_k) + \mathbf{w}_k^{(i)} \tag{6.32}$$

Measurement Update Step: The ensemble prediction is refined using the latest measurements, forming a comprehensive understanding of the current state:

$$\mathbf{z}_k^{(i)} = \mathbf{h}(\mathbf{x}_k^{(i)}) + \mathbf{v}_k^{(i)} \tag{6.33}$$

Statistical Integration: The ensemble's mean and covariance are calculated and used to compute the Kalman gain, which is then used to update the state estimates:

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{6.34}$$

$$\mathbf{x}_k^{(i)} = \mathbf{x}_k^{(i)} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k^{(i)})) \tag{6.35}$$

Ensemble Transformation: Following the update, the ensemble is transformed to maintain consistency with the underlying probability distribution:

$$\mathbf{x}_k^{(i)} = \mathbf{T}_k \mathbf{x}_k^{(i)} \tag{6.36}$$

where $\mathbf{T}_k$ is a transformation matrix computed from the ensemble covariance, ensuring that the updated ensemble accurately reflects the updated beliefs about the system's state after incorporating new measurements.

Using this advanced EnKF methodology, the system effectively captures nonlinear rotational dynamics and high-precision positional tracking. The seamless adjustment to sudden head movements is achieved through sophisticated quaternion propagation within the ensemble framework, demonstrating the filter's robustness and adaptiveness in dynamic AR environments.

6.2.7 Advanced Auxiliary Particle Filter (APF) for Dynamic Head Tracking in AR Flight Simulators

The Auxiliary Particle Filter (APF), is a sequential Monte Carlo method, is particularly adept at handling nonlinear and non-Gaussian systems through a set of particles representing the state distribution. This adaptation utilizes an auxiliary variable approach to enhance sampling efficiency, crucial in dynamic, noisy environments such as AR flight simulation. This implementation is informed by foundational techniques described in seminal works such as those by Gordon et al. [230] and is further refined

to integrate quaternion-based orientation updates and RGB-D sensor data for high-precision tracking [231].

State Representation The state vector $\mathbf{x}_k$ for this application encompasses comprehensive motion descriptors:

$$\mathbf{x}_k = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} & q_w & q_x & q_y & q_z \end{bmatrix}^\top \tag{6.37}$$

This vector includes the 3D position, velocity, and quaternion components for orientation, which are critical for accurate motion tracking in three-dimensional space.

Process Model The process model incorporates detailed kinematic and rotational dynamics:

$$\mathbf{x}_k^{(i)} = \begin{bmatrix} \mathbf{r}_k^{(i)} + \mathbf{v}_k^{(i)} \Delta t \\ \mathbf{v}_k^{(i)} + \mathbf{a}_k^{(i)} \Delta t \\ \mathbf{q}_k^{(i)} \otimes \Delta \mathbf{q}^{(i)} \end{bmatrix} + \mathbf{w}_k^{(i)} \tag{6.38}$$

Here, $\mathbf{w}_k^{(i)}$ represents process noise, modelled as Gaussian noise with covariance $\mathbf{Q}_k$, facilitating realistic simulations of head movements [223].

Measurement Model This model leverages advanced sensor data to track head position accurately:

$$\mathbf{z}_k = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{v}_k \tag{6.39}$$

Measurement noise $\mathbf{v}_k$ is also assumed to follow a Gaussian distribution with covariance $\mathbf{R}_k$, ensuring the filter accounts for sensor inaccuracies.

APF Algorithm

Particle Initialization:

$$w_0^{(i)} = \frac{1}{N}, \quad \text{for } i = 1, 2, \ldots, N \tag{6.40}$$

Particles are initialized to represent the initial belief about the state distribution with uniform weights.

Prediction Step: Each particle is advanced according to the process model:

$$\mathbf{x}_k^{(i)} = \mathbf{f}(\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_k) + \mathbf{w}_k^{(i)} \tag{6.41}$$

Update Step: Particle weights are updated based on the likelihood of the new measurement:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \cdot p(\mathbf{z}_k | \mathbf{x}_k^{(i)}) \tag{6.42}$$

Weights are then normalized to ensure they sum to one:

$$\hat{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}} \tag{6.43}$$

Resampling Step: To prevent particle degeneracy, a resampling step is performed:

$$\mathbf{x}_k^{(i)} \sim \{\mathbf{x}_k^{(j)}, \hat{w}_k^{(j)}\}, \quad \text{for } j = 1, 2, \ldots, N \tag{6.44}$$

This step ensures that particles with higher likelihood are more likely to propagate into the next generation, focusing computational resources on the most probable hypotheses.

In this enhanced APF setup, auxiliary variables and an advanced measurement model allow the filter to adapt swiftly to changes in the user's head orientation and position, providing robust and accurate tracking critical for immersive AR experiences.

6.2.8 Advanced Adaptive Fusion Filters for Precise Head Tracking in Flight Simulators

In flight simulation, achieving highly accurate and responsive head tracking is paramount for an immersive experience. To address this, we implement Adaptive Fusion Filters that integrate the strengths of both Kalman and Particle Filters. This strategic fusion optimizes accuracy, minimizes latency, and offers enhanced robustness against sensor noise and the dynamic uncertainties inherent in real-time simulation environments.

Fusion Step: The core of our adaptive fusion approach involves a sophisticated weighting mechanism that combines the estimates from Kalman and Particle Filters to produce a single, optimized state estimation:

$$\begin{aligned} \text{FusedPosition} = {} & \text{kalmanWeight} \cdot \text{kalmanFilteredPosition} \\ & + \text{particleWeight} \cdot \text{particleFilteredPosition} \end{aligned} \tag{6.45}$$

$$\begin{aligned} \text{FusedOrientation} = {} & \text{Slerp}(\text{kalmanFilteredOrientation}, \\ & \text{particleFilteredOrientation}, \text{particleWeight}) \end{aligned} \tag{6.46}$$

Key elements in this fusion process include:

- kalmanWeight and particleWeight are dynamically determined based on each filter's performance, evaluated through normalized innovation rates or confidence measures of each filter's output accuracy.
- Slerp$(\cdot)$ or Spherical Linear Interpolation is utilized for blending quaternion-based orientations, ensuring a smooth transition between orientations provided by the individual filters.

Dynamic Weighting: Weights assigned to each filter's outputs are computed adaptively:

$$\text{kalmanWeight} + \text{particleWeight} = 1 \tag{6.47}$$

These weights are adjusted continuously based on real-time analysis of the normalized innovation, which measures the deviation of actual measurements from the predicted states. Such dynamic weighting allows the fusion process to be highly responsive to changes in sensor noise levels and the dynamic characteristics of the tracked object (i.e., the user's head in the simulator).

Implementation and Impact: Implementing this fusion strategy allows for a robust, flexible response to varying conditions within the flight simulation, ensuring that head tracking remains precise and aligned with the user's movements. By adapting the outputs of Kalman and Particle Filters, the system maintains high accuracy and minimal response time, which is crucial for maintaining realism and immersion in AR-based flight simulators.

This fusion approach not only capitalizes on the individual strengths of each filtering technique but also mitigates their respective weaknesses, particularly in environments where quick adjustments to new information are critical. The dynamic adjustment of weights based on performance metrics ensures that the fusion filter continually optimizes its performance, enhancing the overall reliability and effectiveness of the head tracking system.

6.2.9 Experimental Setup

This experiment evaluates the head-tracking performance of various filters in AR-based flight simulation, focusing on critical performance metrics: accuracy, latency, and jitter. The primary objective is to identify the filtering algorithm offering the highest accuracy and lowest latency, as benchmarked against prior AR-tracking standards.

We recruited ten participants (five male, five female) aged 20–40, all with normal vision and no prior experience in flight simulation. Participants were briefed on the study objectives and provided informed consent. Each participant used a Microsoft HoloLens 2 headset, calibrated to track head movement within the flight simulator accurately. The simulation included random noise and movement patterns to mimic real-world head-tracking scenarios.

Experimental Design: Leveraging the advanced head-tracking features of Microsoft HoloLens 2, we developed a Unity 3D Engine application to render a virtual cube in the user's view. While the cube model is a simplification of real cockpit scenarios, it was selected as a visual proxy to allow both users and researchers to clearly perceive key tracking behaviors—such as latency, jitter, and drift—during filter operation. These perceptual cues are essential for evaluating head-tracking performance, especially when obtaining direct ground truth in complex, real-world AR flight scenarios is challenging. The cube thus serves as a surrogate for more complex visual stimuli, enabling controlled, repeatable testing of the filter's responsiveness and accuracy under representative motion patterns.

This AR-based setup, enhanced with integrated Kalman and particle filter libraries, enabled high-fidelity data collection on head movements. The Unity 3D environment was equipped with the MixedRealityToolkit for essential functionalities, a MixedRealityPlayspace for the camera, and a HeadTrackingManager for tracking logic and data logging. User instructions were presented through a UICanvas, while the virtual cube mirrored the participant's head movement in real-time. Figure 6.2 provides an illustration of the Unity 3D scene and headset setup.

Participants began the session by focusing on the virtual cube. As they moved their heads, the HoloLens sensors continuously recorded head position and orientation, yielding data for real-time filter analysis. The tracking system employed Kalman and particle filters with parameters chosen through rigorous literature review and

FIGURE 6.2: Experimental Setup and Cockpit Instrumentation. The figure shows the flight simulator setup used for head-tracking studies. Panels (a-b) display the pilot's view and calibrated cockpit, (c) the subject interacting with the cockpit, and (d-e) the visual stimuli and response interface. This setup enables precise tracking and analysis of head movements in response to simulated flight conditions.

empirical testing, aimed at balancing accuracy and computational efficiency. Table 6.1 summarizes the primary parameters implemented in each filter.

6.3 Results

6.3.1 Overview of Filter Performance

This section presents a quantitative evaluation of six filtering techniques and their fusion counterparts for head tracking in an AR flight environment. The filters compared include Extended Kalman Filter (EKF), Sequential Importance Resampling (SIR), Unscented Kalman Filter (UKF), Unscented Particle Filter (UPF), Ensemble Kalman Filter (EnKF), and Adaptive Particle Filter (APF), along with their fused counterparts.

Table 6.2 summarizes the performance metrics—Position RMSE, Rotation RMSE, Average Latency, and Variance Latency—across all evaluated filters. These metrics enable the assessment of tracking accuracy, computational efficiency, and stability for each method.

TABLE 6.1: The table summarizes the key parameters used in various filters implemented in the experiment application.

Symbol	**Value**	**Justification**
Q	0.001f	Represents the default process noise covariance. Chosen to handle small process noise for state prediction.
R	0.01f	Represents the default measurement noise covariance. Selected to accommodate minor measurement noise.
P	1.0f	Represents the initial error covariance. Chosen to start with a moderate level of uncertainty.
CLAMP_MAX_D	10.0f	Represents the maximum clamping distance. Used to limit the maximum prediction error.
adaptiveFactor	0.3f	Adaptive factor for noise adjustment. Helps in providing a dynamic adjustment based on the application requirement.
numParticles	200	Number of particles used in particle filtering. Balances computational cost and accuracy.
latencyCompensation	0.01f	Compensation factor for latency in the system. Helps in synchronizing real-time tracking with the system response.
alpha	1e-3f	Spread of the sigma points in Unscented Kalman Filter. Chosen for stable sigma point distribution.
beta	2.0f	Parameter for the Unscented Kalman Filter. Optimized for Gaussian noise.
kappa	1.0f	Secondary scaling parameter in Unscented Kalman Filter. Chosen to balance scaling of state points.
numEnsembles	200	Number of ensembles in Ensemble Kalman Filter. Ensures a robust state estimate.
measurementSize	3	Dimension of the measurement vector. Based on the head-tracked data points.
stateSize	4	Dimension of the state vector. Includes position and orientation.
ukfProcessNoise	0.001f	Process noise for Unscented Kalman Filter. Set as default.
ukfMeasurementNoise	0.01f	Measurement noise for Unscented Kalman Filter. Set as default.
upfProcessNoise	0.001f	Process noise for Unscented Particle Filter. Set as default.
upfMeasurementNoise	0.01f	Measurement noise for Unscented Particle Filter. Set as default.

6.3.2 Filter Performance Analysis for AR Flight Simulation Head Tracking

The performance metrics of the tested filters are detailed in Figures 6.3 (Position RMSE), 6.4 (Rotation RMSE), and 6.5 (Latency). These figures illustrate each filter's accuracy and computational efficiency, which are critical for AR-based flight simulations.

Extended Kalman Filter (EKF): EKF achieved a low positional RMSE (0.054460m) and rotational RMSE (0.004262m) with minimal latency (0.0055ms). Its innovation-based adaptive estimation, which incorporates velocity and acceleration in its state vector, enabled it to handle nonlinear dynamics effectively. These characteristics make it particularly suitable for real-time applications requiring rapid updates, such as

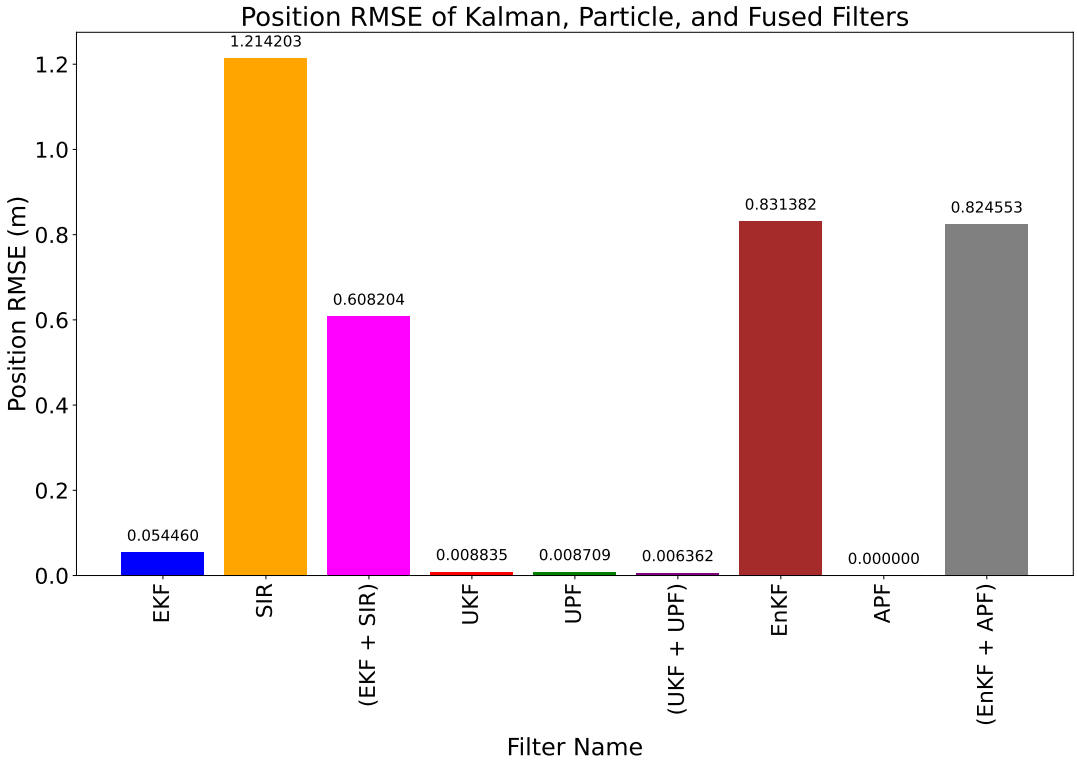


FIGURE 6.3: Position RMSE of various filters. APF shows the lowest RMSE (highest accuracy), while SIR shows the highest RMSE (lowest accuracy). Among fused filters, UKF + UPF perform best.

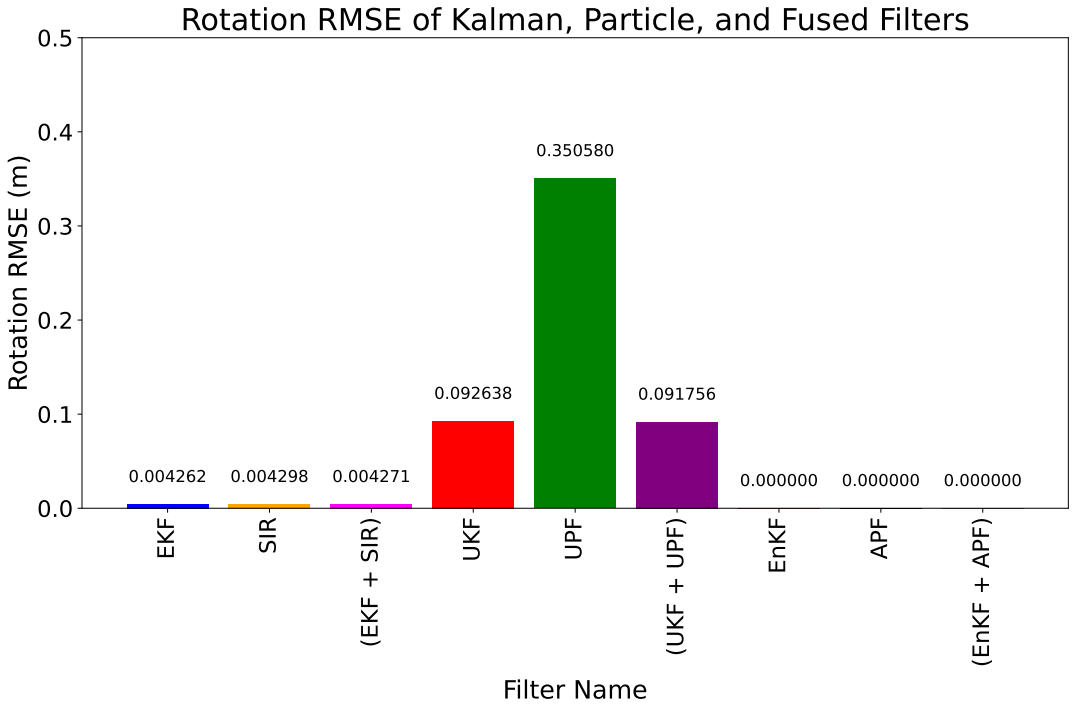


FIGURE 6.4: Rotation RMSE of Kalman, Particle, and Fused Filters. EKF, SIR, and EKF + SIR show low RMSE, indicating high accuracy. UKF and UPF have higher RMSE, with UPF being the highest. UKF + UPF performs well among fused filters, while EnKF, APF, and EnKF + APF achieve zero RMSE.

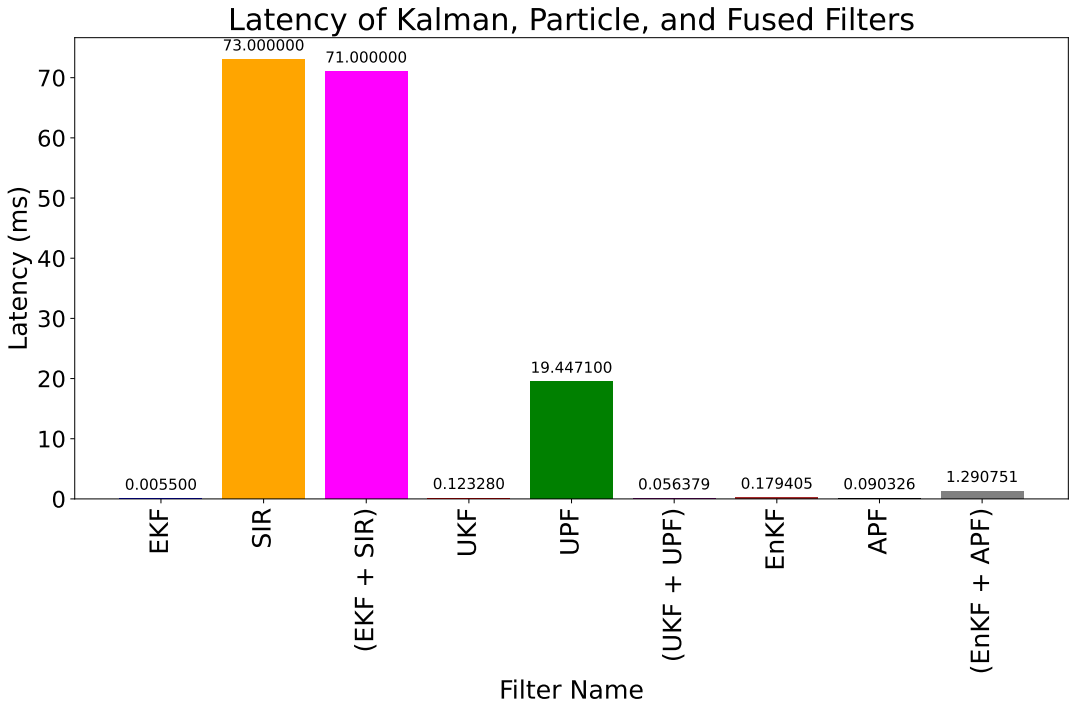


FIGURE 6.5: Latency of Kalman, Particle, and Fused Filters. EKF shows the lowest latency, while SIR and EKF + SIR have the highest latencies. UPF has a moderate latency, with the fused filter UKF + UPF performing well among the fused options.

TABLE 6.2: Performance Comparison of Filtering Techniques for AR-based Head Tracking

S/N	Filter	Position RMSE (m)	Rotation RMSE (m)	Latency (ms)	Variance Latency (ms)
1	EKF	0.0545	0.0043	0.0055	0.0145
2	SIR	1.2142	0.0043	73.0	80.2
3	Fused (EKF + SIR)	0.6082	0.0043	71.0	63.1
4	UKF	0.0088	0.0926	0.1233	0.0066
5	UPF	0.0087	0.3506	19.4	0.9910
6	Fused (UKF + UPF)	0.0064	0.0918	0.0564	0.0003
7	EnKF	0.8314	0.0000	0.1794	0.0279
8	APF	0.0000	0.0000	0.0903	0.0095
9	Fused (EnKF + APF)	0.8246	0.0000	1.2908	0.4001

immersive AR training or flight simulations.

Sequential Importance Resampling (SIR): SIR demonstrated robustness in managing non-linearities, achieving a positional RMSE of 1.214203m and a rotational RMSE of 0.004298m. However, its high latency (73.00ms) limits its applicability for time-sensitive AR applications.

Unscented Kalman Filter (UKF): UKF performed robustly under non-linear conditions, with low positional RMSE (0.008835m), moderate rotational RMSE (0.092638m), and latency of 0.123280ms, making it a strong candidate for dynamic scenarios requiring prompt responses.

Unscented Particle Filter (UPF): UPF showed strength in noisy environments, achieving a positional RMSE of 0.008709m and a rotational RMSE of 0.350580m. While latency was moderate (19.44710ms), UPF's resilience to non-Gaussian noise highlights its utility for complex AR simulations.

Ensemble Kalman Filter (EnKF): EnKF achieved a positional RMSE of 0.831382m and zero rotational error, with a latency of 0.179405ms. Its ability to handle complex datasets makes it suitable for high-precision applications.

Adaptive Particle Filter (APF): APF demonstrated exceptional accuracy with near-zero positional and rotational RMSE and low latency (0.090326ms), making it ideal for real-time AR tasks requiring precision.

Fusion Methods: To address the limitations of individual filters, fusion methods were explored to enhance accuracy and balance latency trade-offs:

EKF + SIR Fusion: This fusion improved positional RMSE to 0.608204m and rotational RMSE to 0.004271m. However, latency remained high (71ms), limiting its applicability to scenarios where accuracy is prioritized over responsiveness.

UKF + UPF Fusion: This fusion achieved the best balance of accuracy and efficiency, reducing positional RMSE to 0.006362m and rotational RMSE to 0.091756m, with a latency of 0.0564ms. It is particularly suitable for dynamic, real-time AR simulations.

EnKF + APF Fusion: While this fusion achieved a positional RMSE of 0.8246m and zero rotational error, it exhibited increased latency (1.2908ms), indicating its utility in high-accuracy applications that are not latency-critical.

These findings underscore the importance of aligning filter selection with specific application requirements. The results demonstrate that fusing UKF and UPF for position tracking and EnKF and APF for rotation effectively validates our approach by achieving high tracking accuracy and reduced latency. This makes the proposed

method particularly well-suited for immersive training in augmented reality (AR) environments, with a particular emphasis on applications such as flight simulations.

6.4 Discussion

The results highlight the trade-offs among accuracy, latency, and computational efficiency across the evaluated filters, influencing their suitability for specific AR applications. EKF's low positional and rotational RMSE, combined with its minimal latency, positions it as a strong candidate for real-time applications requiring quick and precise updates. Its innovation-based adaptive estimation and ability to handle non-linear dynamics make it particularly valuable for immersive training and flight simulation scenarios.

In contrast, SIR demonstrated robustness in managing non-linearities but suffered from significant latency. While this limits its applicability in time-sensitive contexts, its ability to handle complex non-linear models suggests potential utility in less time-critical applications.

Similarly, UPF excelled in noisy environments but exhibited moderate latency, making it suitable for tasks where noise resilience is critical, such as dynamic or complex AR scenarios. Fusion methods provided a means of overcoming individual filter limitations. For instance, the EKF + SIR fusion balanced accuracy and robustness but was constrained by high latency. The UKF + UPF fusion, by combining UKF's stability with UPF's noise resilience, offered an optimal balance between precision and responsiveness, making it ideal for dynamic AR simulations. Meanwhile, although highly accurate, the EnKF + APF fusion introduced latency challenges that may limit its use in real-time applications.

Beyond performance metrics, the AKPF algorithm's adaptability underscores its potential for a broad range of AR applications. The algorithm can address computational constraints while maintaining essential tracking accuracy by tuning key parameters, such as particle count and noise covariance. This flexibility makes AKPF suitable for various AR use cases, from resource-constrained mobile devices to high-performance headsets used in training and navigation.

6.4.1 Adaptability Across Diverse AR Applications

The Adaptive Kalman-Particle Filter (AKPF) algorithm developed in this study demonstrates significant potential for deployment across diverse augmented reality (AR) environments, each characterized by unique head-tracking requirements. For instance, medical AR systems designed for surgical navigation demand stable and precise tracking to accommodate fine, deliberate head movements [232]. Conversely, gaming applications often involve rapid and unpredictable head turns, necessitating robust tracking solutions to handle highly dynamic movement profiles [233]. The AKPF's tunable parameters—particle count, process noise, and measurement noise covariance—allow it to adapt effectively to these varied demands. For example, the algorithm can prioritize stability and precision in medical contexts by refining noise covariance settings, while optimizing particle count to enhance responsiveness for fast-paced

gaming scenarios. Furthermore, the flexible design of the AKPF ensures compatibility with hardware devices of varying computational capacities. On resource-limited mobile AR platforms, reducing the particle count or adjusting noise parameters can maintain algorithmic efficiency without significantly compromising tracking accuracy. In contrast, high-performance AR devices, such as dedicated headsets for training simulations or automotive heads-up displays, can leverage increased particle counts and refined noise adjustments to achieve enhanced tracking precision [234]. This adaptability positions the AKPF as a versatile and robust solution for various AR applications spanning entertainment, healthcare, automotive, and other industries. Its capacity to balance tracking accuracy with computational efficiency establishes its utility across contexts with varying hardware constraints and movement demands.

6.4.2 Real-Time Performance and Computational Feasibility

Performance testing was conducted on Microsoft HoloLens 2 to evaluate the AKPF's latency, precision, and accuracy under dynamic conditions. This balance is achieved by adaptive parameter tuning—particle count, process noise (Q), and measurement noise covariance (R)—in response to real-time motion dynamics. For instance, adjusting particle count in low-noise environments or fine-tuning noise parameters for applications with high-motion variability enables the filter to accommodate various head movement patterns and noise conditions, making it suitable for diverse AR tasks, as outlined in Table I in the appendix.

Unlike conventional implementations that rely on static parameters, our approach adaptively tunes particle count and noise parameters in response to real-time motion dynamics. This responsiveness is critical for achieving both computational feasibility and tracking robustness on resource-constrained AR hardware like the HoloLens 2.

Monte Carlo simulations [235], incorporating randomized sensor noise and head movement dynamics, validated the filter's robustness across 100 runs. Gaussian noise with a standard deviation of 0.001 to 0.01 was modelled to replicate real-world conditions, ensuring realistic performance evaluations. These noise levels were selected through exploratory testing and iterative calibration. The noise model captured AR environment variability, accounting for both small fluctuations and significant deviations in sensor readings. This confirmed AKPF's adaptability under variable sensor conditions and dynamic movement profiles.

While the AKPF demonstrates strong real-time accuracy and robustness, its use of particle filtering introduces a computational burden that increases with particle count. Higher particle counts provide better tracking precision but also lead to increased latency and processing load. This can impact overall system responsiveness, especially during prolonged high-dynamic motion or in resource-limited environments. Table I in the appendix and Figure 6.6 illustrate that filters incorporating particle components tend to have higher latency compared to pure Kalman filters, underscoring the trade-off between accuracy and computational cost. To mitigate this, AKPF uses an adaptive strategy that dynamically allocates computational effort. Under stable tracking conditions, the system relies more on the Kalman filter, which is computationally lightweight. During rapid or non-linear movements, particle filters are selectively engaged to preserve accuracy, thereby reducing average resource usage while maintaining robustness.

Although the current evaluation was performed on HoloLens 2, a high-performance AR device, the AKPF algorithm is designed with scalability in mind. Techniques such as reducing particle count, optimizing resampling steps, and adjusting update frequency enable deployment on lower-power devices. Preliminary profiling suggests that real-time performance can be maintained with fewer particles and simplified update logic. This implies that AKPF can operate efficiently on mobile AR platforms such as smartphones and lightweight AR glasses. Testing and profiling on such platforms are recommended to evaluate its generalizability further and fine-tune AKPF parameters for optimal performance across a broader spectrum of AR devices.

6.4.3 User Experience Evaluation: Motion Sickness Potential

Motion sickness (or cybersickness) significantly impacts user experience in AR applications, particularly in flight simulation. Research indicates that latency, jitter, and tracking inaccuracies contribute to symptoms such as nausea, dizziness, and disorientation [207].

This study assesses motion sickness potential by examining key factors:

Latency: High latency, the delay between head movements and corresponding AR rendering, increases the risk of motion sickness by causing sensory conflicts.

Jitter: latency variation results in unstable tracking, leading to unexpected virtual movements that exacerbate user discomfort.

Accuracy: Precise tracking ensures alignment between real and virtual movements, reducing sensory discrepancies and associated symptoms.

Our results indicate that high-latency filters, such as SIR for position tracking and UPF for rotation, elevate the potential for motion sickness due to pronounced response lag (Figure 6.6). The SIR filter's extensive position error range indicates high jitter, increasing motion sickness risks. In contrast, fused filters like UKF + UPF and EnKF + APF demonstrate low latency and consistent tracking, as reflected by reduced error spread in both position and rotation. This stability promotes a more synchronous experience between physical and virtual cues, mitigating the likelihood of motion sickness.

For flight simulations, where users often perform rapid and dynamic head movements, minimizing latency and jitter is critical to avoid disorientation and nausea. The fusion methods evaluated in this study effectively address these challenges, supporting smoother and more responsive tracking.

Our fusion method's ability to minimize latency and enhance accuracy directly supports an improved user experience by reducing the likelihood of motion sickness, making it an ideal solution for high-performance AR-based applications. Although this study emphasizes motion sickness as a critical usability factor, future evaluations could explore additional user experience metrics, such as visual comfort, ease of interaction, and long-term usability.

6.4.4 Comparative Analysis of AR Flight Simulation Head Tracking

This section delves into the comparative analysis of individual filters and fusion methods tailored for AR-based head tracking in flight simulations, with a focus on the

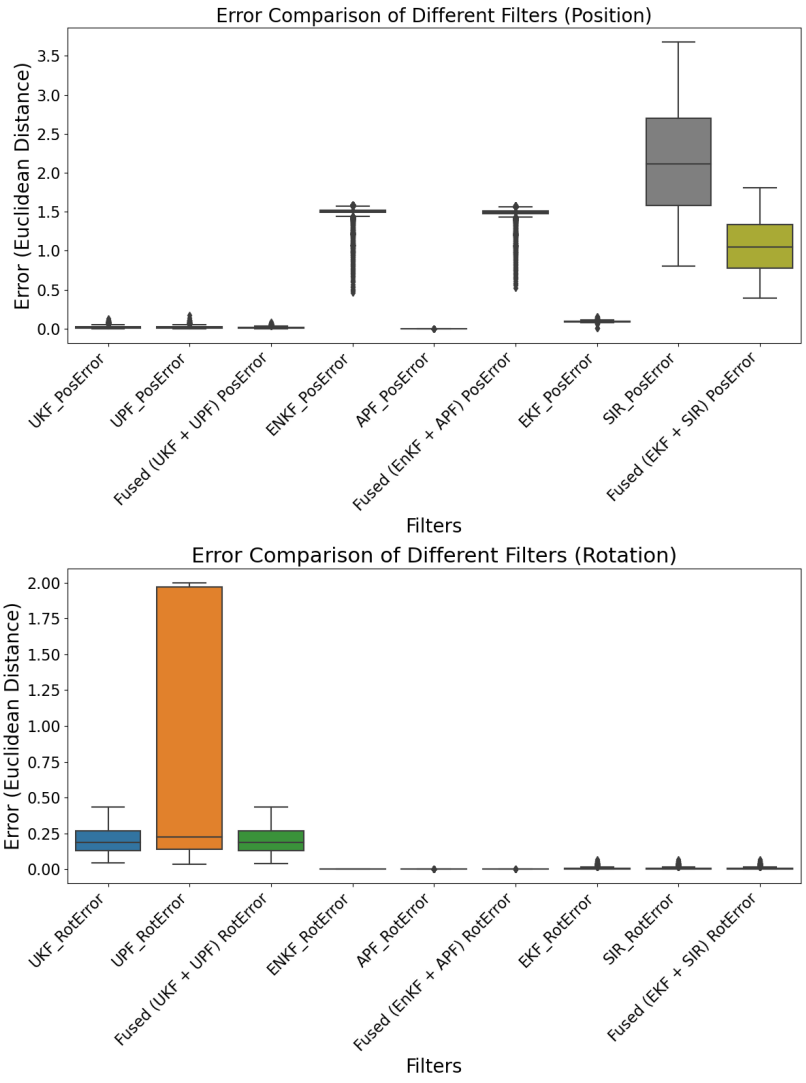


FIGURE 6.6: Error Comparison of Filters for Position and Rotation. The top plot shows position errors, with SIR having the highest variability, while fused filters like UKF + UPF (position) and EnKF + APF (rotation) show lower errors. The bottom plot highlights rotation errors, with UPF showing the highest and fused filters demonstrating better consistency and accuracy.

accuracy of position and rotation metrics, measured as the Root Mean Square Error (RMSE), over time. A sophisticated noise-reduction approach has been utilized to refine the estimation of actual positional data by compensating for the raw head-tracking data with the most accurate filter predictions. This method, commonly applied in advanced signal processing, significantly enhances position accuracy by mitigating inherent noise in the data collection process [236].

EKF, SIR, and EKF-SIR Fusion

The RMSE data for the Extended Kalman Filter (EKF), Sequential Importance Resampling (SIR), and their fusion (EKF-SIR) are meticulously illustrated in Figures 6.7 and 6.8. The EKF is distinguished by its stable and low RMSE values, affirming its suitability for real-time applications where latency is a critical factor. Although the SIR filter excels in handling non-linear contexts, its higher RMSE and associated latency somewhat constrain its utility in highly dynamic scenarios. Conversely, the EKF-SIR fusion method demonstrates a synergistic balance, achieving superior accuracy and stability compared to using the SIR filter alone.

The efficacy of the EKF in this AR application is consistent with findings from other significant studies, such as those by Manzano [237], which have highlighted the filter's exceptional performance in 3D tracking contexts and its robust capability to manage non-Gaussian measurements. These characteristics make the EKF an invaluable component of advanced tracking systems in AR, particularly within the demanding environments of flight simulation.

UKF, UPF, and UKF-UPF Fusion

Further analysis is conducted on the Unscented Kalman Filter (UKF), the Unscented Particle Filter (UPF), and their fusion, which are depicted in Figures 6.9 and 6.10. The UKF consistently shows stable, low RMSE, rendering it practical for real-time tracking applications. While the UPF effectively manages non-linearities, it does so at the cost of slightly higher RMSE and increased latency. The fusion of UKF and UPF integrates the stability of the UKF with the non-linear handling capabilities of the UPF, offering an optimized balance of accuracy and robustness across diverse tracking tasks.

Research such as that by Luo [238] and [239] supports these outcomes by demonstrating how enhancements in UPF techniques bolster tracking performance under dynamic conditions, while robust applications of UKF maintain consistent performance standards in AR settings. This alignment of findings underscores the potential of filter fusion techniques to enhance the flexibility and applicability of AR technologies across various domains.

EnKF, APF, and EnKF-APF Fusion

Figures 6.11 and 6.12 illustrate the performance of EnKF, APF, and their fusion. While EnKF handles large datasets effectively, its accuracy diminishes under dynamic conditions. APF achieves consistently low RMSE across both position and rotation, balancing accuracy with moderate latency. The EnKF-APF fusion combines these strengths but exhibits increased latency, making it suitable for tasks prioritizing accuracy over responsiveness.

In evaluating the performance of AR head-tracking filters, particularly the fused filters, it is essential to consider both frame count and time count metrics. Frame count analysis allows us to assess the filter's precision and stability across individual data frames. This is particularly significant in AR applications where high-frequency

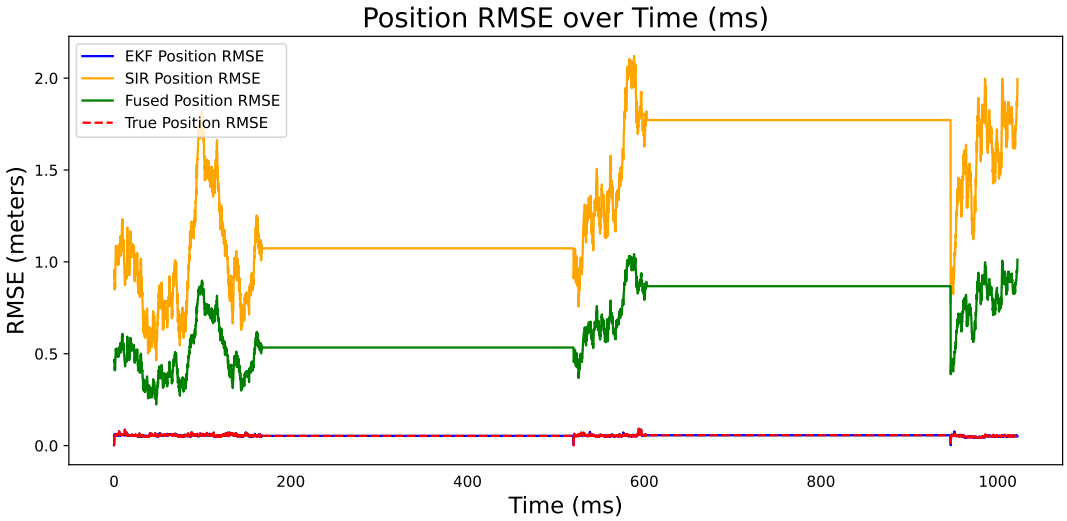


FIGURE 6.7: Root Mean Square Error (RMSE) of position estimates over time for EKF, SIR, and fused EKF-SIR filters.

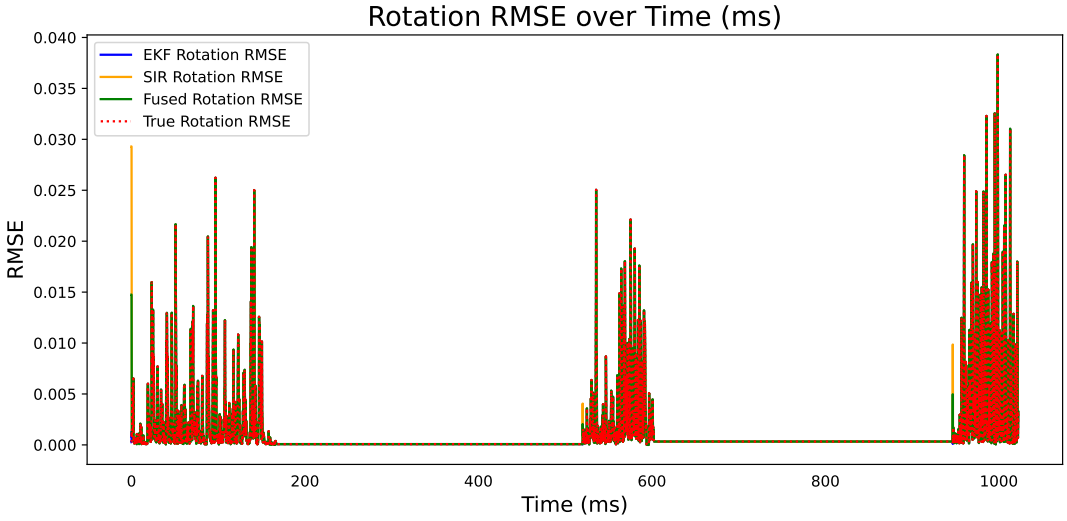


FIGURE 6.8: Root Mean Square Error (RMSE) of rotation estimates over time for EKF, SIR, and fused EKF-SIR filters.

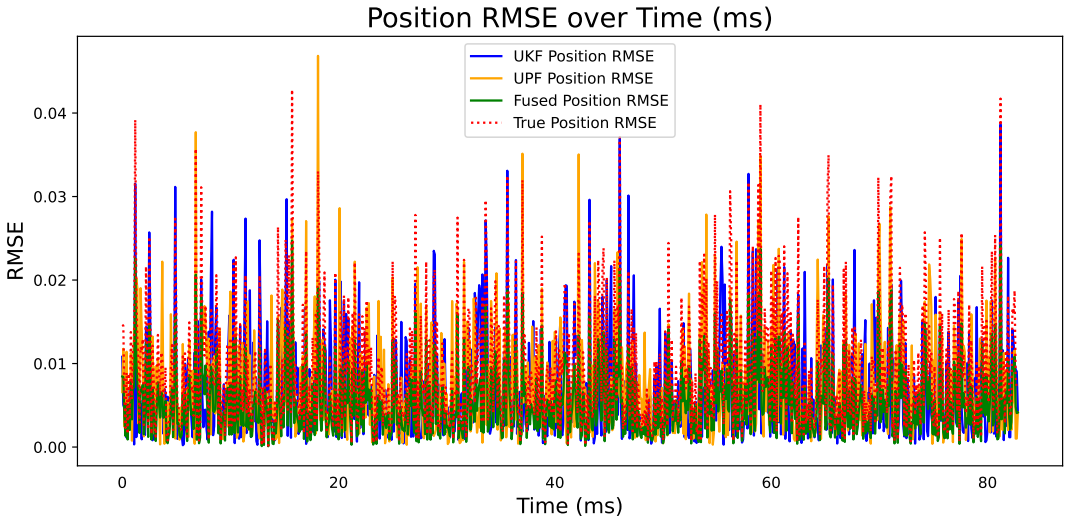


FIGURE 6.9: Position RMSE over time for UKF, UPF, and fused UKF-UPF filters in AR head tracking.

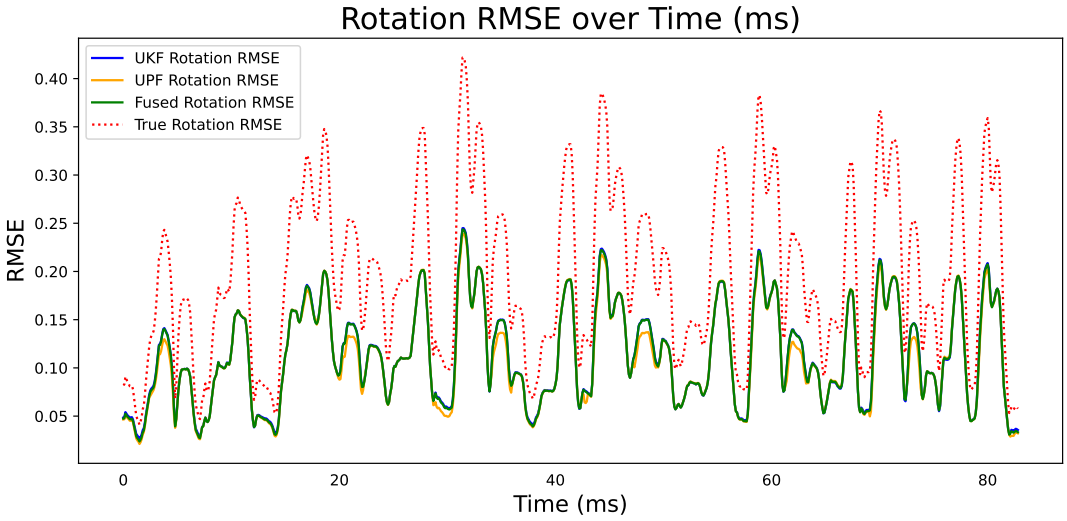


FIGURE 6.10: Rotation RMSE over time for UKF, UPF, and fused UKF-UPF filters in AR head tracking.

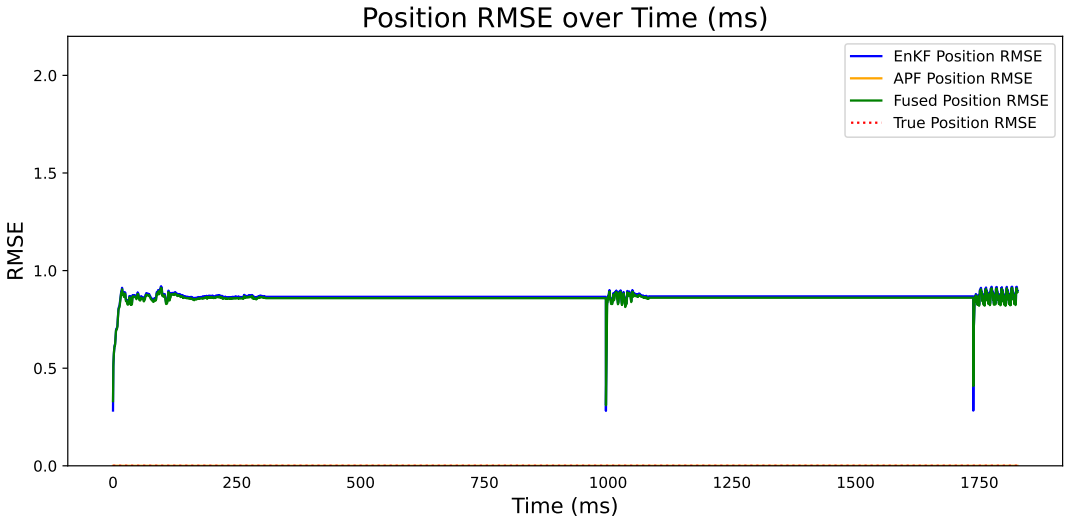


FIGURE 6.11: Position RMSE over time for EnKF, APF, and fused EnKF-APF filters in AR head tracking.

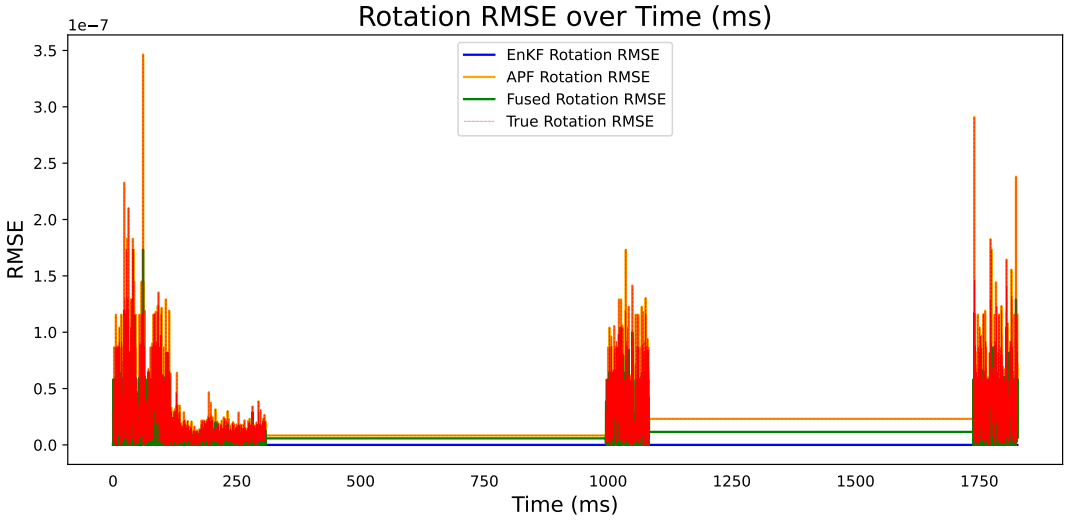


FIGURE 6.12: Rotation RMSE over time for EnKF, APF, and fused EnKF-APF filters in AR head tracking.

updates are crucial for maintaining an immersive experience. The RMSE trends analyzed over frame counts, as shown in Figures 6.13, illustrate how the fused filters manage to maintain reduced error margins consistently across extensive operations, highlighting their robustness and reliability in dynamic environments.

Conversely, time count analysis sheds light on the temporal efficiency of these filters. This metric is crucial for understanding the filters' performance in real-world conditions where processing speed and computational overhead can impact the overall system effectiveness. The fused filters' ability to sustain low RMSE over extended time periods, as depicted in these graphical analyses, underscores their suitability for long-duration AR applications, where both accuracy and performance consistency over time are paramount.

Trajectory Analysis

Figures 6.14 through 6.19 present a series of 3D trajectory plots, offering a comparative visualization of actual versus filtered trajectories for various filters and fusion methods. These plots illustrate the effectiveness of different filtering techniques in addressing noise and improving trajectory estimation. Specifically, the figures include comparisons across Extended Kalman Filters (EKF), Unscented Kalman Filters (UKF), Ensemble Kalman Filters (EnKF), Sequential Importance Resampling (SIR), Unscented Particle Filters (UPF), Auxiliary Particle Filters (APF), and fusion-based approaches.

The actual trajectories, which are inherently noisy due to sensor inaccuracies and environmental disturbances, serve as a baseline for evaluating the performance of the filters. By contrast, the filtered trajectories reveal how each method mitigates noise and improves path stability. The fusion methods, in particular, combine the strengths of individual filters to deliver trajectories that balance accuracy and stability, which is essential for real-time applications such as augmented reality (AR).

The elements of the trajectory analysis are summarized below:

Actual Trajectory (Red Dotted Line): This representation of the unfiltered, noise-affected data highlights the inherent limitations of raw sensor measurements. The significant deviations caused by noise underscore the critical need for effective filtering methods to ensure reliable trajectory estimation.

Filtered Trajectories Using EKF, UKF, and EnKF (Blue Line): The Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Ensemble Kalman Filter (EnKF) demonstrate substantial noise reduction capabilities. These methods stabilize the trajectory across the X, Y, and Z coordinates, ensuring smoother paths. Their performance is particularly effective under linear or mildly non-linear conditions where the assumptions of Gaussian noise hold true.

Filtered Trajectories Using SIR, UPF, and APF (Orange Dashed Line): These particle-filter-based approaches handle probabilistic smoothing under highly non-linear and non-Gaussian scenarios. Their ability to adapt to complex dynamics is reflected in the smoother trajectories they produce. The APF often demonstrates superior performance among these due to its improved resampling strategies.

Fused Trajectories (Green Line): Fusion-based trajectories leverage the individual strengths of the aforementioned filters. These methods optimise trajectory accuracy and stability by intelligently combining the outputs of multiple filters. This approach

FIGURE 6.13: Comparison of RMSE values for Kalman, Particle, and Fused Filters across three datasets: EKF_SIR, UKF_UPF, and EnKF_APF, showcasing their performance and accuracy over increasing frame counts.

is particularly advantageous for AR environments, where achieving a balance between high accuracy and computational responsiveness is crucial.

Figure 6.14: This figure depicts the 3D trajectory of true and filtered positions obtained from the EKF, SIR, and fused methods across the X, Y, and Z coordinates. The visual comparison emphasizes the enhanced stability and accuracy achieved through filtering and fusion techniques.

Figure 6.15: Similar to Figure 6.14, this plot focuses on rotational trajectories. It demonstrates the effectiveness of EKF, SIR, and fusion-based methods in reducing rotational noise and stabilizing angular trajectories.

Figure 6.16: This figure presents the 3D trajectory of true and filtered positions using UKF, UPF, and fused methods. It illustrates the benefits of advanced Kalman and particle filters in accurately capturing positional dynamics.

Figure 6.17: The corresponding rotational trajectories for UKF, UPF, and fused methods are shown here. The fusion approach is particularly effective in stabilizing rotations compared to individual filters.

Figure 6.18: This plot highlights the 3D trajectory of true and filtered positions using EnKF, APF, and fused methods. The EnKF and APF are shown to effectively address positional inaccuracies, while the fusion method further optimizes trajectory performance.

Figure 6.19: The final figure focuses on rotational trajectories filtered using EnKF, APF, and fused methods. It underscores the ability of fusion approaches to provide enhanced stability in both angular and positional estimates.

This analysis underscores the importance of effective filtering and fusion techniques in trajectory estimation. While individual filters exhibit specific strengths, fusion methods combine these strengths to achieve a balanced performance. This is particularly relevant for AR and other real-time applications, where trajectory accuracy and system responsiveness are paramount. The visualizations provided in Figures 6.14 through 6.19 provide clear evidence of the advancements made possible through modern filtering and fusion methodologies.

6.5 Conclusion

This chapter has presented the development and performance evaluation of an Adaptive Kalman-Particle Filter (AKPF) framework designed to enhance head tracking fidelity in augmented reality (AR)-based flight simulation environments. This hybrid approach offers a low-latency solution well-suited for high-demand AR applications, such as flight simulation, where accuracy and responsiveness are critical.

Through a comprehensive series of simulations and experimental validations, the AKPF demonstrated superior performance across key tracking metrics, including position and rotation accuracy, root mean square error (RMSE), and latency. Comparative analysis with conventional filters—such as the Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Ensemble Kalman Filter (EnKF), Sequential Importance Resampling (SIR), and Auxiliary Particle Filter (APF)—highlighted the AKPF's effectiveness in maintaining low-latency tracking with minimal drift and high precision, particularly under dynamic and high-load cockpit scenarios.

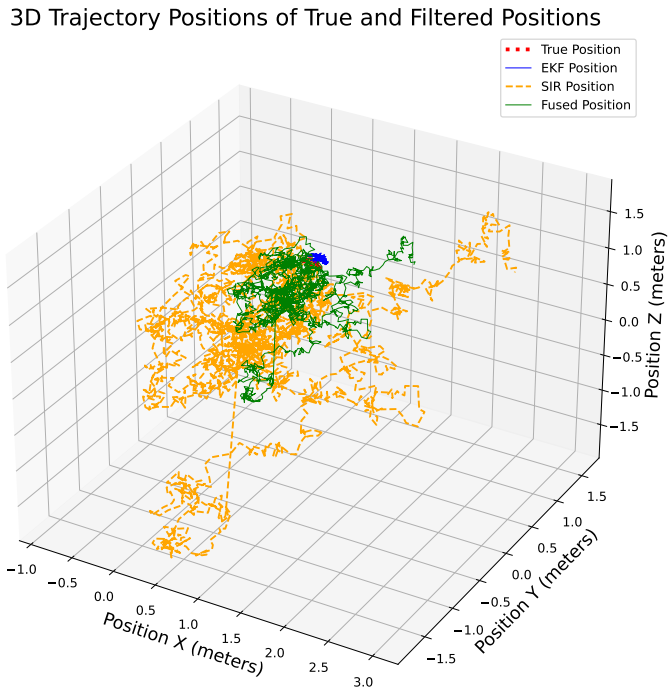


FIGURE 6.14: 3D trajectory of true and filtered positions (EKF, SIR, and Fused) across X, Y, and Z coordinates.

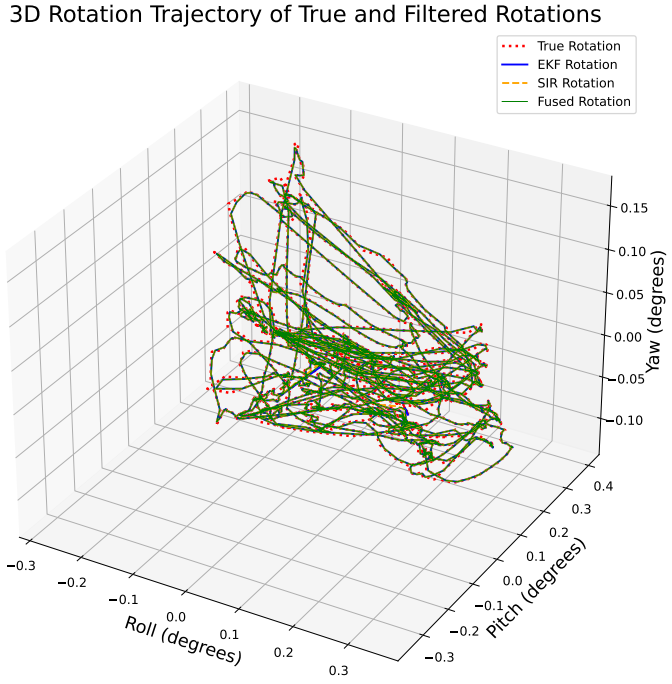


FIGURE 6.15: 3D trajectory of true and filtered rotations (EKF, SIR, and Fused) across X, Y, and Z coordinates.

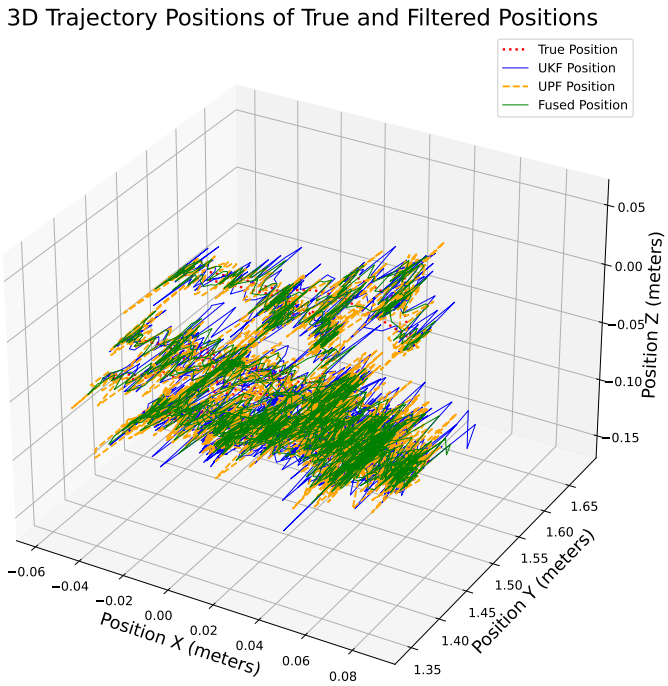


FIGURE 6.16: 3D Trajectory Plot of True and Filtered Positions (UKF, UPF, and Fused) to X, Y, and Z coordinates

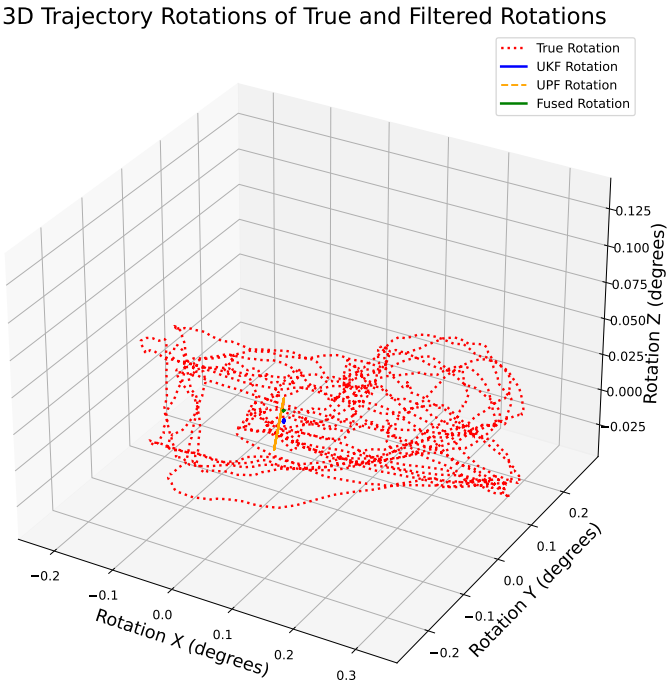


FIGURE 6.17: 3D Trajectory Plot of True and Filtered Rotation (UKF, UPF, and Fused) to X, Y, and Z coordinates

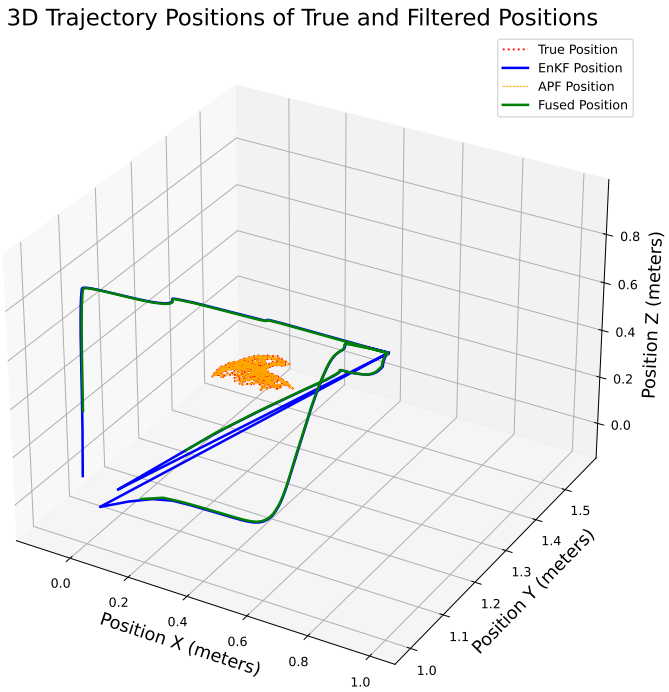


FIGURE 6.18: 3D Trajectory Plot of True and Filtered Positions (EnKf, APF, and Fused) to X, Y, and Z coordinates

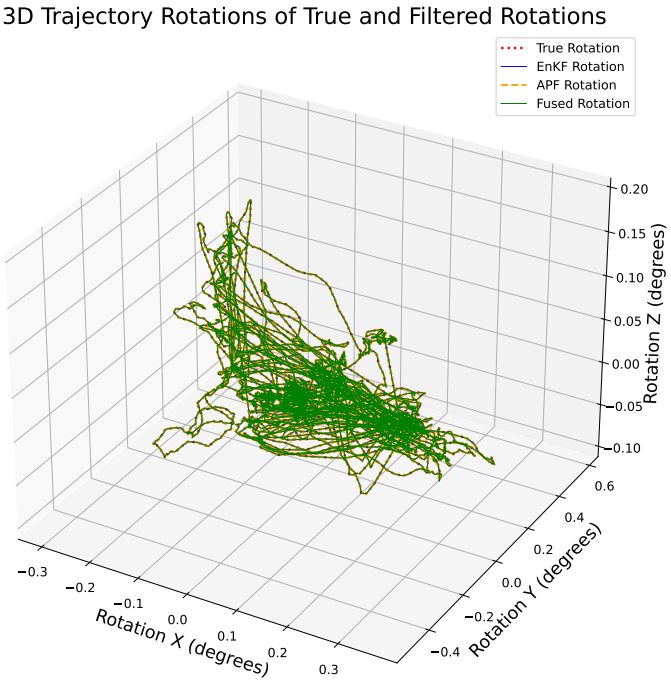


FIGURE 6.19: 3D Trajectory Plot of True and Filtered Rotation (EnKf, APF, and Fused) to X, Y, and Z coordinates

The fusion approach notably enhanced the temporal consistency and spatial stability of head pose estimates, which are critical for maintaining precise alignment between virtual overlays and physical cockpit components. The reduced error margins achieved by the AKPF contribute significantly to the realism and reliability of AR-enhanced training simulations, supporting the core objectives of immersive pilot education and operational readiness. The outcomes of this chapter affirm that the adaptive filtering strategy provides a scalable and computationally efficient solution for real-time head tracking in AR systems. By dynamically adjusting to sensor noise and motion variance, the AKPF enables reliable performance even under variable environmental conditions, such as changes in cockpit lighting or pilot head movement patterns. Pilots can engage in longer training sessions with minimised risk of motion sickness.

Future work will focus on extending this approach to encompass multimodal sensor fusion, including LiDAR and depth-based data, and exploring its applicability to full-body motion tracking and outdoor AR environments. The AKPF framework also holds promise for cross-domain adaptation in other AR-intensive fields, such as surgical training, industrial maintenance, and remote collaboration systems. This chapter makes a significant contribution to AR tracking methodologies, providing a foundation for the continued development of high-fidelity, adaptive, and low-latency filtering systems in next-generation AR simulation platforms.

7 User Identification in Augmented Reality (AR) via Deep Learning and Multimodal Interaction

7.1 Introduction

Augmented Reality (AR) headsets are increasingly being integrated into flight simulator training environments to provide immersive, interactive learning experiences for pilots [240]. Devices such as the Microsoft HoloLens 2 can overlay virtual cockpit instruments, navigation aids, and other flight data onto a physical simulator, enhancing situational awareness and procedural training for student pilots. However, as AR systems become common in aviation training, new challenges emerge in multi-user and resource-constrained settings. In training academies, AR headsets are often shared among different trainees, and multiple pilots or instructors may collaborate in a shared augmented environment. This raises a critical need for reliable user identification within the AR simulator: the system must know which trainee is using the device at any given time to securely load the correct user profile, apply personalized settings, and maintain the continuity of each user's presence (avatar) in collaborative simulations. In other words, maintaining secure, personalized, and continuous authentication for each trainee pilot has become essential to realize the benefits of AR-enhanced flight training fully. Traditional login methods or manual profile selection are cumbersome in an immersive cockpit scenario and prone to error [241], particularly when rapid transitions between users are required. A more seamless solution is required to ensure that shared AR devices instantly recognize each trainee and allow them to continue their training where they left off [242].

To address these challenges, this chapter introduces a behavioural biometric identification system designed explicitly for AR-based flight simulation. The goal is to enable personalized and secure interactions in the augmented cockpit by implicitly recognizing the user through their natural movements and behaviours, rather than relying on external credentials or interruptions. The proposed solution leverages the Simultaneous Localization and Mapping (SLAM)–enabled sensors of a HoloLens 2 headset to profile the trainee unobtrusively. In particular, the framework continuously collects multimodal data streams, including the pilot's gaze trajectories, head movement patterns, and hand gesture dynamics, during routine training tasks and interactions with the AR content. These rich behavioral signals serve as a unique "Signature" for each user in the system [242], [243].

A multi-branch Long Short-Term Memory (LSTM) neural network is then employed to learn and recognize the temporal behavioural signature of each pilot. By

training on sequences of AR interaction data, the LSTM model captures idiosyncratic patterns (for example, how a particular pilot scans instruments (eye gaze), moves their head (head movement), and makes hand gestures while executing a checklist) that distinguish one trainee from another [244]. This deep learning approach enables the system to continuously and passively verify the user's identity, eliminating the need for explicit login steps, biometric scans, or calibration from the trainee once the system is in use.

Integrating this behavioral biometric system into the AR flight simulator yields several key benefits for aviation training. First, it provides a form of continuous authentication, the headset actively confirms that the current user is indeed the authorized trainee throughout the session [245]. This enhances security and prevents scenarios where an unapproved user could accidentally (or intentionally) use another pilot's AR device or access their training data [246]. Second, it enables a high degree of personalization: once identified, the trainee's individualized settings (e.g., display configurations, lesson progress, or adapted difficulty level) can be automatically loaded, creating a seamless training experience that adjusts to the individual [247], [248]. Third, the system ensures user continuity in both single-user and multi-user (collaborative) simulation sessions. In a single-user context, if a headset is handed from one student to another or if a trainee returns for a new session, the AR system swiftly recognizes the change in identity and transitions to the appropriate profile without disruption. In a multi-user collaborative scenario, for example, two trainees and an instructor participating in a networked AR flight exercise, the identity-aware system maintains consistent user–avatar mappings, so that each virtual avatar or identifier in the shared augmented space reliably corresponds to the correct real-world pilot [249]. This continuity is crucial for realism and coordination: instructors and fellow trainees must trust that the actions and performance data attributed to a given virtual avatar indeed belong to the intended person. By solving the identity ambiguity on shared AR platforms, the training environment can prevent confusion and preserve the integrity of performance assessments and interactive tasks.

The use of head movement tracking data as part of the biometric feature set means this approach aligns closely with the broader aims of the dissertation. Earlier chapters focused on improving head-tracking accuracy and low-latency spatial mapping in AR (using SLAM) to ensure that virtual overlays remain stable and realistic within the cockpit. In this context, the same AR tracking capabilities are redirected toward a new yet complementary purpose: determining the identity of the individual being tracked. In essence, this chapter extends the SLAM-based head and gaze tracking innovations from prior chapters beyond pure navigation and visualization, into the realm of user-specific identification [250]. This extension underscores how advanced tracking not only anchors content to the cockpit with high fidelity but can also anchor the experience to the correct user. By coupling the spatial tracking of the AR headset with a deep learning model for user recognition, the system contributes to the dissertation's goal of creating adaptive, intelligent flight simulator environments. The AR training platform becomes not only spatially aware (knowing where the trainee is looking or moving) but also identity-aware (knowing who the trainee is), which is a key step toward fully personalized pilot training.

This chapter presents a novel behavioural biometric profiling system for AR flight

simulators that enhances training security and personalization through continuous, implicit user authentication. By integrating identity recognition with AR head-tracking, the system aims to enhance the fidelity and safety of flight simulation training, ensuring that each trainee's immersive experience is securely tailored to them alone, even in complex shared-device or multi-user training scenarios.

The contributions of this chapter are summarised as follows:

- We develop a novel behavioral biometric identification framework for AR-based flight simulator training, leveraging head movement, gaze, and hand motion data from the HoloLens 2 to passively recognize users in both single-user and collaborative environments.
- We design and implement a deep multi-branch LSTM model that learns temporal signatures from these multimodal sensor streams, enabling robust, real-time, privacy-preserving, and calibration-free authentication that enhances personalization, security, and identity continuity.
- We extend prior spatial tracking developments in the dissertation by repurposing head and gaze data not only for rendering but also for secure, identity-aware simulation, demonstrating the complementary benefits of integrating identity recognition into AR training systems.

7.1.1 Related Work

Simulation-based training has long been central to aviation education, and numerous studies affirm its effectiveness in developing pilot skills [35]. Modern research is increasingly integrating immersive technologies, such as augmented and virtual Reality (AR/VR), into flight simulators to enhance realism and training outcomes. For instance, AR interfaces can overlay instrument cues and scenario augmentations in the cockpit, which has been shown to reduce pilot workload and improve situational awareness [203]. Similarly, VR-based simulators provide fully synthetic environments for safe yet authentic practice of flight operations [251]. These advances underscore that immersive simulation is not only viable but advantageous in aviation training [12]. However, one under-explored aspect in prior simulator research is the personalization of training. Current AR/VR flight trainers typically assume the user's identity is known (e.g., via manual login) and then apply generic training content. This stands in contrast to evidence that individualized training adjustments can benefit learning outcomes [247], [248]. Enabling the system to recognize the trainee at run-time would pave the way for truly adaptive, pilot-specific training regimens. This vision has motivated recent interest in behavioural biometric profiling within immersive environments, where unique patterns in a user's interactions are leveraged to continuously and unobtrusively identify or authenticate them [241]. In the context of flight simulation, such identity-aware technology would ensure that each pilot's progress and custom settings seamlessly carry over between training sessions, even when using shared AR equipment [241], [242].

Gaze-Based Identification: The way users move their eyes can serve as a biometric identifier. Research has shown that specific eye movements metrics, such as saccadic speed, fixations, and blink patterns, are reliably person-specific [252]. Eye tracking

is now available on modern HMDs (e.g., HoloLens 2, HTC Vive Pro Eye), enabling gaze-based authentication techniques. These approaches can be categorized into two types: explicit and implicit. In explicit gaze-based authentication, users perform predetermined eye movements, also known as "gaze gestures," as a form of authentication. Notable examples include systems where users input a secret pattern by moving their eyes (such as EyePass and blink-pattern passwords) [253]. While explicit gaze schemes can be effective, they require active user participation and memorization of a pattern. In contrast, implicit gaze-based methods monitor the user's natural eye behaviour during normal AR/VR usage without requiring any deliberate action [254]. Such implicit systems continuously validate identity based on unconscious gaze traits, offering seamless security. For instance, Rigas et al.[255] utilized features such as saccade vigour and acceleration, achieving reasonably good accuracy in distinguishing users via eye movements. Overall, gaze tracking has emerged as a viable biometric modality; however, gaze alone can sometimes be influenced by contextual factors (such as lighting and visual stimuli), and thus, it is often combined with other signals for greater robustness.

Head Movement and Pose Another behavioral biometric is the user's head movement pattern. Wearing an AR or VR headset, each person naturally has a unique way of orienting their head and responding to stimuli. Prior work leveraged this by designing specific head-motion challenges: Wang and Zhang [256] asked users to perform a head-nodding task in VR and then used the resulting motion data to identify individuals. Similarly, Mustafa et al. [257] demonstrated that even simple head-orientation data from a Google Cardboard (a smartphone-based VR sensor) could differentiate users to a significant extent. Beyond task-specific motions, continuous head-tracking data can be used for implicit identification. Sivasamy et al. [258] developed a method that utilizes only inertial head movements from an HMD and reported very high identification accuracy (over 99% in specific controlled scenarios) by analyzing features such as movement direction changes and velocity during natural behaviours. These results suggest that head kinematics carry identifiable signatures. The HoloLens 2's built-in IMU and inside-out tracking cameras provide a rich stream of such head pose data that can be mined for biometric patterns. A key consideration with head-motion biometrics is stability over time: factors such as fatigue or posture changes can introduce variability, so robust algorithms must distinguish between persistent traits (e.g., reaction speeds, movement smoothness) and transient behaviour.

Hand and Gesture Dynamics: In AR, users interact with virtual content via hand gestures (e.g., HoloLens's air-tap and pinch gestures) and motion controllers in VR. The way a person reaches, points, or gestures in mid-air can also function as a biometric identifier. Olade et al. [259] introduced a swipe-based authentication in VR inspired by touchscreen patterns, showing that the trajectory of a user's hand swipe in 3D space can be used to distinguish them, albeit with some performance cost compared to simpler 2D swipes. Other studies have employed more complex interaction tasks; for instance, one approach required users to perform virtual object manipulation tasks (such as throwing a ball or archery in a game) and extract features from their hand and arm movements. Liebers et al. [243] found that using such VR tasks (e.g., virtual bowling and archery games) to gather biometric data improved identification accuracy, especially when combined with normalization for each user's physical arm length

and height. These controlled tasks provide rich motion data; however, designing task-specific biometrics may not always be practical in real applications. An ideal system would recognize users during ordinary interactions without requiring additional actions.

Despite promising results in prior studies, several gaps remain in the literature. Many studies focus on a single modality or require users to perform a designated task, which may not accurately reflect operational use in training scenarios. There is a need for implicit, continuous authentication methods that function during arbitrary AR interactions without interrupting the user. Achieving cross-session reliability remains an open challenge—models trained on one session often deteriorate when the same user returns on a different day or engages in different activities, indicating possible overfitting to session-specific characteristics. Addressing this requires robust feature learning, potentially through sensor fusion or adaptive algorithms, to capture enduring personal traits rather than transient behaviours.

Relatively few works have explored the complete fusion of multimodal data in realistic AR settings. Bhalla et al. [260] is a notable example targeting AR, as it continuously authenticated HoloLens users by analyzing head IMU signals and interaction data, achieving approximately 92.7% balanced accuracy on a small set of five users. While encouraging, that study was limited in scope and did not incorporate data from the eyes or hands. To date, no published work has combined eye gaze, head motion, and hand gesture biometrics on a HoloLens 2 for real-time user identification in an immersive flight simulation context.

This doctoral research directly addresses these gaps by proposing and evaluating a novel multimodal biometric framework that fuses gaze, head motion, and hand gesture data for continuous, implicit pilot identification in AR-based training. As detailed in the following chapters, the framework is designed to operate in real-time, maintain cross-session robustness, and integrate seamlessly into immersive simulation environments, enabling truly personalized pilot training.

7.2 Methodology

7.2.1 Experimental Setup and User Interaction Design

The interaction system was developed in Unity (version 2021 or later) using C# and deployed on the Microsoft HoloLens 2 mixed-reality headset. The system supports three primary modalities of interaction: gaze-based eye tracking, head-tracked manipulation, and gesture-driven hand interaction. Each of these modalities operates independently and can be explicitly activated by the user. This modular design enables controlled and isolated data acquisition, facilitating multimodal behavioural analysis and biometric profiling.

To systematically collect user behavior data across these interaction types, we implemented a unified event-logging and control flow, described in Algorithm 3. The system continuously captures gaze fixation events, head-controlled object manipulation, and natural hand gestures during each task. Each modality contributes a time-synchronised stream of sensor data, which is later processed and fused for behavioural modelling.

Algorithm 3 Multimodal Data Acquisition and Profiling

Require: Unity scene with eye-, head-, and hand-tracking modules
Ensure: Structured event log and behavioral embeddings
Initialize session with *userId* and *name*
Present task-selection UI
while timeRemaining **do**
  if eyeTrackingActive **then**
    for each $sphere \in virtualSpheres$ **do**
      if gazeDuration$(sphere) > 0.5\,\text{s}$ **then**
        Trigger visual feedback on *sphere*
        Log GAZEENTER and GAZEEXIT with timestamps and positions
      end if
    end for
  end if
  if headTrackingActive **then**
    if fixatingOn(*cylinder*) **then**
      Attach *cylinder* to head (compute offset)
      Log HEADPICK
    end if
    if reachesTargetPlane(*cylinder*) **then**
      Log HEADPLACE and advance target
    end if
  end if
  if handGestureActive **then**
    if $pinchDetected \vee tapDetected$ **then**
      Log HANDGESTURE with timestamp and object metadata
    end if
  end if
end while
Synchronize gaze, head, and hand streams to a common timeline
Form fixed-length sequences and feed to LSTM for identity classification

A total of 55 participants (35 male, 20 female, ages 18–40) took part in the study. From their interactions across the three modalities—eye gaze, head tracking, and hand gestures—we collected 10,032 synchronized behavior sequences, each 3 seconds long. All participants had varied experiences with augmented reality technologies. Before the experiment, each participant was briefed on the objectives and procedures and provided their informed consent. All interactions were conducted using the Microsoft HoloLens 2 headset.

The experimental session began with the system prompting users to enter their name and a unique user ID. This information was used to associate all interaction data with the corresponding individual, allowing for structured storage and later analysis. Once registered, the participant was guided to a main menu displayed within the augmented reality environment. This menu presented options for the three interaction tasks—eye tracking, head tracking, and hand gesture exercises—accompanied by brief on-screen instructions. The participants could choose the order in which they completed the tasks and were given a total of 10 minutes to explore all three exercises (see Fig. 7.1 for an overview of the scene layout). A real-time countdown timer remained visible throughout the session, providing continuous feedback on the remaining time.

Eye-Tracking Interaction. In the eye-tracking task, participants were immersed in a virtual 3D space populated with spherical objects evenly distributed throughout their

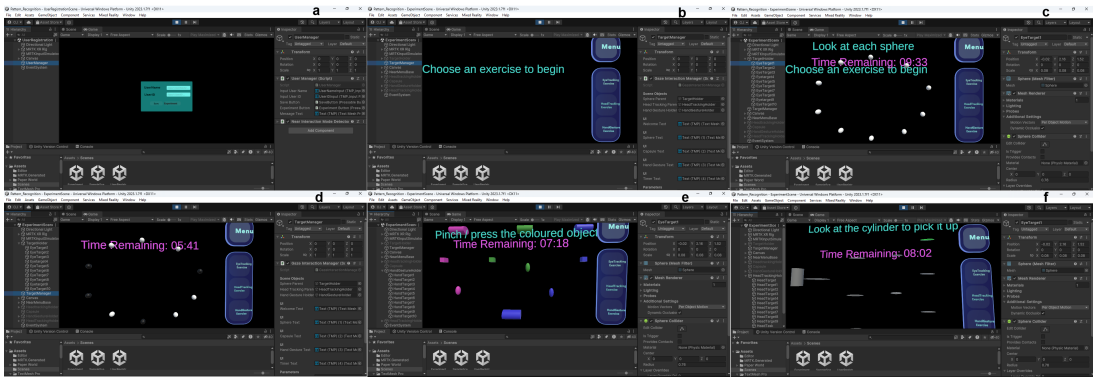


FIGURE 7.1: User interface and interaction flow of the experimental system developed in Unity for the HoloLens 2. It illustrates (a) user registration, (b) task selection menu, (c-d) eye-tracking task, (e) hand gesture task, and (f) head-tracking task with real-time countdown feedback.

visual field. Each sphere was designed to respond to user attention. When a participant fixated on a sphere, the system measured the duration of gaze using continuous raycasting from the user's forward head direction. If the gaze were maintained for more than 0.5 seconds, the sphere would visually respond by gradually changing its colour to red. Simultaneously, it would turn black, providing visual confirmation of successful interaction. Each instance of gaze initiation and termination was logged, along with timestamps and the 3D position of both the sphere and the user's head orientation at the time of interaction.

Head-Tracking Interaction: For the head-tracking task, participants interacted with a virtual cylinder object and a series of spatial target planes arranged within the environment. The task began when a participant fixated on the cylinder, triggering it to attach virtually to their head via a computed spatial offset. From that point on, the cylinder moved in sync with the user's head, maintaining a fixed relative position and orientation. The user's goal was to move the cylinder into proximity to a highlighted target plane. When the cylinder reached a pre-defined spatial threshold near the target (approximately 2.35 units), the system logged a placement event, and the next target in the sequence became active. This interaction sequence was repeated, encouraging precise and controlled head movements. Each placement was recorded along with spatial and temporal data to assess head-based motor control and alignment accuracy.

Hand-Gesture Interaction: In the gesture-based interaction task, users engaged with virtual objects using natural hand gestures, such as pinching and tapping. These gestures were recognized in real-time through the HoloLens' hand-tracking system, which monitored the joint positions of both the left and right hands. Objects configured to respond to gesture inputs provided immediate visual feedback, such as changing color upon interaction. Each gesture event was logged with a timestamp, hand identification, and the object being manipulated. This task emphasized naturalistic hand interaction and provided data for analyzing the frequency, duration, and hand dominance of gestures.

Interaction Logging and Feature Capture: Throughout all tasks, user interactions were continuously monitored and recorded in a structured format. Interaction events

were logged to a CSV file named *InteractionEvents.csv* in real-time. Each recorded entry included metadata about the type of interaction (e.g., gaze fixation, head-based placement, or hand gesture), the object involved, and an interaction index. Temporal attributes, such as start and end times, interaction duration, and time between consecutive events, were captured. Spatial properties—including the 3D positions of the head, hands, and virtual objects—were recorded alongside motion-related features such as head and hand velocities. Orientation data was stored using rotational quaternions to represent head and hand pose in 3D space.

Additional derived features, such as the distance between gaze direction and object centre, total object movement, and a categorical estimate of movement speed (slow, medium, or fast), were also computed. A burst logging mechanism was employed to capture high-frequency data surrounding each interaction event. This mechanism recorded data at 5 Hz for the first 30% of a 1.5-second window, then at 2.5 Hz thereafter. This design ensured high temporal resolution during the most critical phases of user motion, allowing for detailed analysis of interaction dynamics and behavior transitions.

Evaluation Strategy for Behavioral Continuity: To evaluate the continuity and generalizability of the trained model, we performed a second phase of experimentation involving a distinct subset of users. These participants performed the same set of interaction tasks (eye gaze, head motion, and hand gestures) as described in the data collection protocol. However, their data were excluded entirely from the training set, ensuring no overlap with the sequences used for model optimization.

The purpose of this evaluation was to simulate a real-world deployment scenario, where a system trained on known users must correctly identify or authenticate new users based solely on previously learned behavioural patterns. The behavioural data of each user was preprocessed into time-aligned sequences and passed through the trained LSTM model for identity prediction.

This strategy allowed us to test the model's capacity for temporal and cross-subject generalization, a critical requirement for any biometric system aiming to support continuous, session-independent authentication. The accuracy of these unseen user sequences serves as an indicator of the model's robustness in practical applications where enrollment and interaction conditions may vary over time.

7.2.2 LSTM-Based Temporal Modelling for Behavioral Biometric Identification

Following the data acquisition pipeline (Algorithm 1), the raw eye gaze, head motion, and hand gesture streams are first synchronized and segmented into fixed-length windows (3-second sequences at 5 Hz, i.e., $T = 15$ time steps). Each sequence is then fed into a multi-branch LSTM network (Figure 7.2) for user identity profiling. We employ recurrent neural networks with long and short-term memory (LSTM) [261] to capture the temporal dependencies and continuity in user behavior patterns. The gating mechanism of the LSTM (input, forget and output gates) [262] enables the model to selectively retain relevant information over time and filter out transient noise, which is crucial to model subtle long-range cues that characterize each user. This long-range temporal modeling capability allows the learned behavioral signature to persist over time and even between different usage sessions. Moreover, using three parallel LSTM branches (one per modality) capitalizes on the complementary nature of

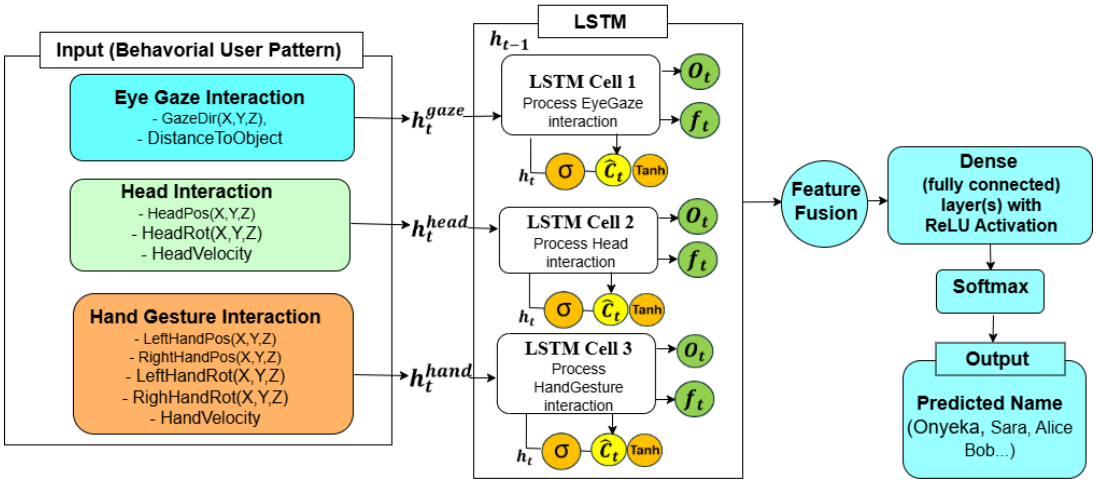


FIGURE 7.2: Overview of the proposed LSTM-based multimodal user identification architecture. Sensor data from three behavioral interaction channels—eye gaze, head movement, and hand gestures—are processed independently through LSTM cells to extract temporal features. Each LSTM outputs hidden states o_t, cell states c_t, and forget gates f_t, capturing temporal dynamics per modality. The outputs are fused and passed through fully connected (dense) layers with ReLU activation, followed by a softmax layer for user identity classification.

the modalities-each branch can learn modality-specific dynamics (e.g., eye movement rhythms vs. head motion patterns), while the later fusion yields a richer and more distinctive representation of the user than any single modality alone.

Each modality $m \in \{\text{gaze, head, hand}\}$ is processed by a dedicated LSTM subsystem implemented in a standard many-to-one configuration. In our architecture, each branch consists of two stacked LSTM layers (each with $H = 256$ hidden units), followed by a dropout layer (rate = 0.3) to mitigate overfitting. A summary of the complete model configuration is provided in Table 7.1. For each time step t ($1 \leq t \leq T$) in branch m, the LSTM cell updates are given by the standard gating equations [262]:

$$\begin{aligned}
i_t^{(m)} &= \sigma\big(W_i^{(m)}\, x_t^{(m)} + U_i^{(m)}\, h_{t-1}^{(m)} + b_i^{(m)}\big),\\
f_t^{(m)} &= \sigma\big(W_f^{(m)}\, x_t^{(m)} + U_f^{(m)}\, h_{t-1}^{(m)} + b_f^{(m)}\big),\\
o_t^{(m)} &= \sigma\big(W_o^{(m)}\, x_t^{(m)} + U_o^{(m)}\, h_{t-1}^{(m)} + b_o^{(m)}\big),\\
\tilde{c}_t^{(m)} &= \tanh\big(W_c^{(m)}\, x_t^{(m)} + U_c^{(m)}\, h_{t-1}^{(m)} + b_c^{(m)}\big),\\
c_t^{(m)} &= f_t^{(m)} \odot c_{t-1}^{(m)} \;+\; i_t^{(m)} \odot \tilde{c}_t^{(m)},\\
h_t^{(m)} &= o_t^{(m)} \odot \tanh\big(c_t^{(m)}\big)\,,
\end{aligned} \tag{7.1}$$

where $\sigma(\cdot)$ denotes the sigmoid function in elemental order, $\tanh(\cdot)$ is the hyperbolic tangent and $\odot$ represents elemental multiplication. In these equations, $x_t^{(m)} \in \mathbb{R}^{d_m}$ is the input vector for the characteristic in time t for the modality m. The vectors

TABLE 7.1: Implementation summary of the multi-branch LSTM classifier used for temporal modeling in behavioral biometric identification.

Component	**Specification**
Sequence Length (T)	15 time steps (3 seconds @ 5 Hz)
Input Dimensionality	$d_{\text{gaze}} = 6$, $d_{\text{head}} = 6$, $d_{\text{hand}} = 6$
LSTM Layers per Branch	2 stacked layers
Hidden Units per Layer	256
Dropout (LSTM)	0.3 (between LSTM layers)
Fusion Strategy	Concatenation of hidden states $h_T^{(m)}$
Dense Layer (Post-Fusion)	Fully connected + ReLU
Dropout (Post-Fusion)	0.3
Output Layer	Softmax over $N = 55$ classes
Loss Function	Cross-Entropy
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Epochs	100
Training/Validation Split	80% / 20%
Parameter Count (per branch)	836,407
Framework	PyTorch 2.0.1

$h_t^{(m)} \in \mathbb{R}^H$ and $c_t^{(m)} \in \mathbb{R}^H$ are the hidden state and cell state of the LSTM (with $H = 256$ in our implementation), and W, U, b are the trainable weight matrices and bias terms for the input, recurrent, and cell update connections, respectively. The gating activations $i_t^{(m)}$, $f_t^{(m)}$, and $o_t^{(m)} \in \mathbb{R}^H$ correspond to the input, forget, and output gates, which regulate the information flow at time t. By design, each branch's LSTM learns the temporal structure of its modality stream in isolation (e.g., gaze fixation transitions, head rotation dynamics, hand gesture timing), thereby preserving modality-specific patterns.

At the end of each sequence ($t = T$), each branch yields a modality-specific embedding given by its final hidden state $h_T^{(m)}$. We concatenate the three branch outputs $h_T^{(\text{gaze})}$, $h_T^{(\text{head})}$, and $h_T^{(\text{hand})}$ to form a single fused feature vector (see Figure 2). This joint representation is then passed through a fully connected dense layer with ReLU activation, followed by a dropout layer (rate = 0.3), and finally a softmax output layer, which produces a probability distribution over the N identity classes. In other words, the network outputs an N-dimensional vector of identity probabilities for each input sequence. The full implementation code, including data preprocessing scripts, model definitions, and training routines, is publicly available at our GitHub repository: OJnwobodo.

7.3 Evaluation Metrics

To comprehensively assess the performance of our proposed multimodal biometric identification framework, we adopt standard metrics widely used in behavioral classification tasks involving temporal and biometric data. These metrics capture both overall system performance and class-wise behavior, and are consistent with established evaluation practices in the pattern recognition community.

7.3.1 Accuracy

Accuracy quantifies the proportion of correct predictions over all instances and provides a coarse measure of overall classification performance:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.2}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

7.3.2 Precision, Recall, and F1 Score

To capture user-specific prediction performance, we compute class-wise and macro-averaged versions of the following:

- **Precision** measures the correctness of positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7.3}$$

- **Recall** reflects the model's ability to retrieve all relevant instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7.4}$$

- **F1 Score** is the harmonic mean of precision and recall:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7.5}$$

These metrics are particularly useful for evaluating model robustness on underrepresented or confusable user identities.

7.3.3 Balanced Accuracy

To account for class imbalance in the dataset, we report balanced accuracy—defined as the mean of recall values across all user classes:

$$\text{Balanced Accuracy} = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_i}{TP_i + FN_i} \tag{7.6}$$

where C is the total number of user classes. This metric ensures that each class contributes equally to the performance evaluation, regardless of its sample frequency [263].

7.3.4 Confusion Matrix

To visualize class-specific prediction behavior, we employ a row-normalized confusion matrix. Each entry at position (i, j) represents the proportion of true class-i instances predicted as class-j. High diagonal values indicate strong per-user recognition, while off-diagonal values reveal patterns of misclassification. This analysis is particularly informative in identifying clusters of users with similar behavioral patterns or overlapping motion profiles, which may warrant feature-space refinement.

7.3.5 t–SNE Embedding Visualization

To qualitatively assess the discriminative capacity of our learned user embeddings, we project the 128-dimensional penultimate-layer LSTM output $\mathbf{z} \in \mathbb{R}^{128}$ into a two-dimensional space using t–SNE [264]. We follow the best-practice settings recommended by Wattenberg *et al.* [265], employing a perplexity of 30, automatic learning rate, and PCA-based initialization. Figure 7.7 (Section V) presents a t–SNE plot of embeddings for ten randomly selected users. Distinct and compact user-specific clusters in this latent space corroborate the model's capacity to encode identity-preserving behavioral signatures, consistent with the high diagonal dominance observed in the confusion matrices.

7.3.6 Chance-Level Baseline

For reference, we report the chance-level accuracy corresponding to a random guesser in a multi-class setting:

$$\text{Chance Level} = \frac{1}{C} \tag{7.7}$$

where C is the total number of users. In our 55-user dataset, this yields a baseline of approximately 1.82%. Including this baseline contextualizes our model's predictive performance, particularly in evaluating whether classification accuracy significantly exceeds random assignment [266].

7.4 Result and Discussion

We trained and evaluated the proposed LSTM-based multimodal identification framework on a dataset comprising 10,032 behavioral sequences collected from 55 users interacting with a Microsoft HoloLens 2 headset. Each sequence includes synchronized data from natural gaze, head, and hand movements captured during unstructured AR tasks. To assess system effectiveness, we conducted both intrasession and cross-session experiments, evaluating the model's ability to distinguish users in repeated and novel sessions. These experiments simulate realistic deployment conditions and assess the robustness of user-specific behavioural signatures over time. As illustrated

in Figure 7.3, participants performed interaction tasks within an augmented reality (AR) environment, allowing the system to collect rich, representative behavioral data under naturalistic, ecologically valid conditions.

The core identification model, based on a multibranch LSTM architecture, was trained using synchronized sequences of eye gaze, head rotation, and hand gesture data. Each input sequence contained parallel samples of the three modalities. As shown in Figure 7.4, the model demonstrated rapid convergence, achieving classification accuracy exceeding 99% on both the training and validation sets within the first 100 epochs. Notably, the loss curves for both sets decreased consistently and converged near zero, reflecting effective optimization and a well-calibrated learning process. The near-identical trajectory of the training and validation performance curves further indicates a strong generalization capacity, with no substantial evidence of overfitting throughout the training cycle.

To further assess classification robustness, we visualized a normalized confusion matrix (Figure 7.5) using a randomly sampled subset of ten users. The matrix shows perfect or near-perfect classification across all selected users, with values concentrated along the diagonal. This confirms the model's precision in capturing user-specific behavioral patterns and highlights its ability to discriminate between individuals with minimal confusion.

To evaluate generalization to new recording contexts, we further present the normalized confusion matrix for the cross-session (unseen) scenario in Figure 7.6. Here, each row shows the correct identification rate (diagonal) and distribution of misclassifications for each user when tested on sessions excluded from training. While the diagonal remains dominant for most users, off-diagonal entries indicate some degradation in accuracy due to session variability—yet the model continues to demonstrate substantial user discriminability and robust generalization to previously unseen data.

To visualise the geometry of the learned embedding space, we project the 128-D penultimate-layer vectors to two dimensions using t-SNE. Figure 7.7 shows the resulting distribution for a representative subset of 10 users. Each colour corresponds to a distinct user; points form well-separated clusters, corroborating the high per-user precision observed in the confusion matrices (Figures 7.5, 7.6).

To qualitatively assess inter-user discriminability, we visualized and compared the motion trajectories of two selected users across all three modalities—eye gaze, head movement, and hand gesture (Figure 7.8). Each trajectory is plotted in the X-Y rotation space and augmented with kernel density estimation (KDE) contours to reveal motion concentration zones [267]. Despite following similar AR tasks, the users displayed distinct spatiotemporal patterns, indicating the presence of consistent, user-specific behavioral signatures. Such differences reinforce the discriminative potential of the multimodal framework and underscore the value of incorporating all three modalities in achieving high classification accuracy. These results confirm the model's ability to learn user-specific behavioral signatures and generalize effectively to previously unseen data, both within the same session (intrasession) and across different sessions (cross-session), thereby demonstrating robustness to session variability.

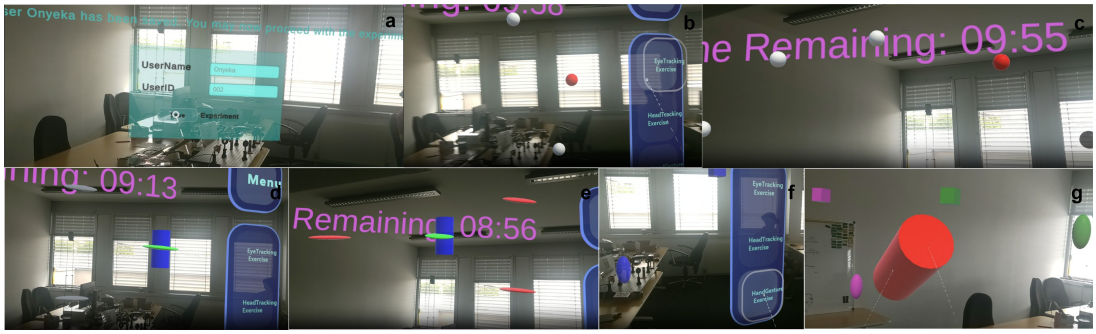


FIGURE 7.3: User interaction in a real-world environment for task-based user identification and data collection. (a) User profile login interface; (b–c) Eye-tracking task; (b, d) Head-tracking task; (e–f) Hand gesture task during human-computer interaction experiment.

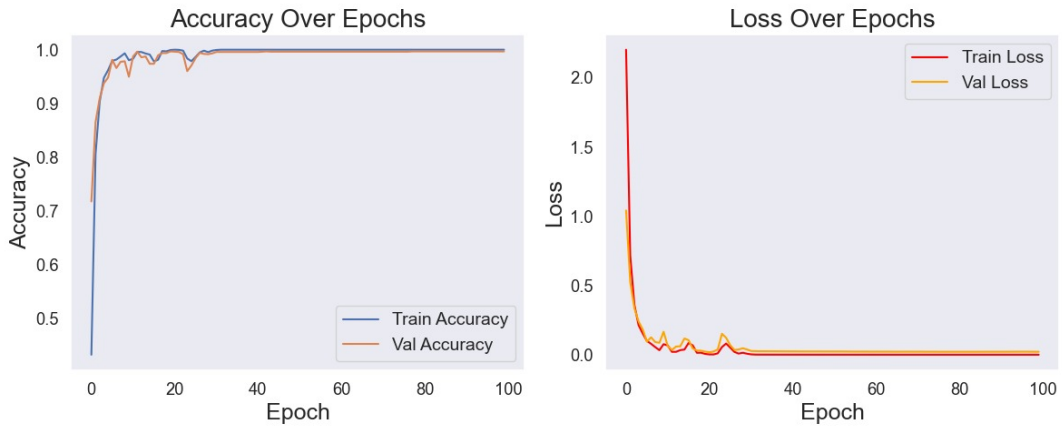


FIGURE 7.4: Training and validation accuracy (left) and loss (right) curves over 100 epochs for the multi-branch LSTM model, showing rapid convergence and minimal overfitting.

7.4.1 Generalization to Unseen Users

To further evaluate the robustness and generalizability of the proposed biometric identification framework, we conducted an in-depth evaluation on unseen users, i.e. participants whose data were withheld from training, to examine open-set performance. simulating a real-world deployment without prior exposure to the test users.

As illustrated in Figure 7.9, we measured the identification accuracy for five such previously unseen users (blue bars), alongside the number of test sequences available for each (green line). The results show a notable range in model performance across these new individuals. The highest classification accuracy reached 93.5%, whereas the lowest was only 4.9%. (For reference, the dashed horizontal line in Figure 7.9 indicates 50% accuracy, equivalent to chance for a binary decision.) Although there is generally a positive correlation between sample count and accuracy, this relationship is not absolute. For instance, User 2 achieved the highest accuracy with over 600 test sequences, suggesting that a larger amount of data supports more reliable

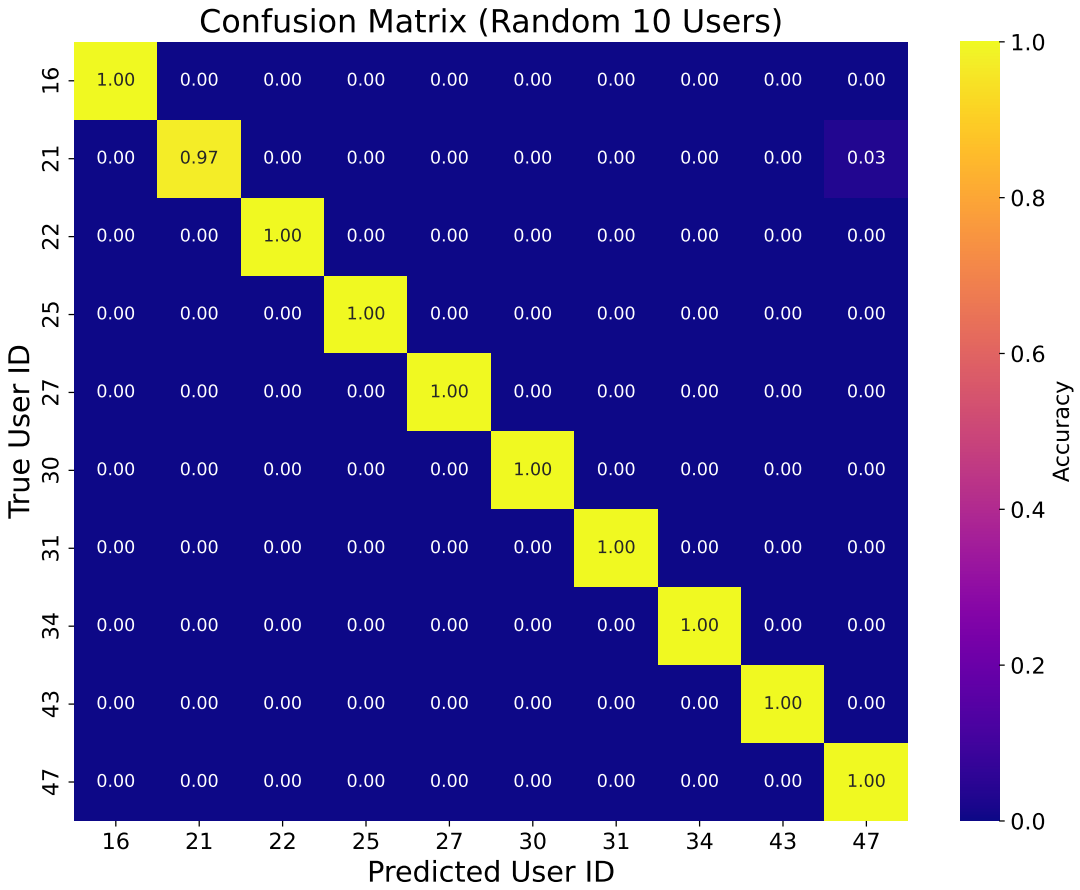


FIGURE 7.5: Normalized confusion matrix for a randomly selected subset of 10 users, demonstrating perfect classification accuracy and high discriminability of user-specific behavioral patterns learned by the multimodal LSTM model.

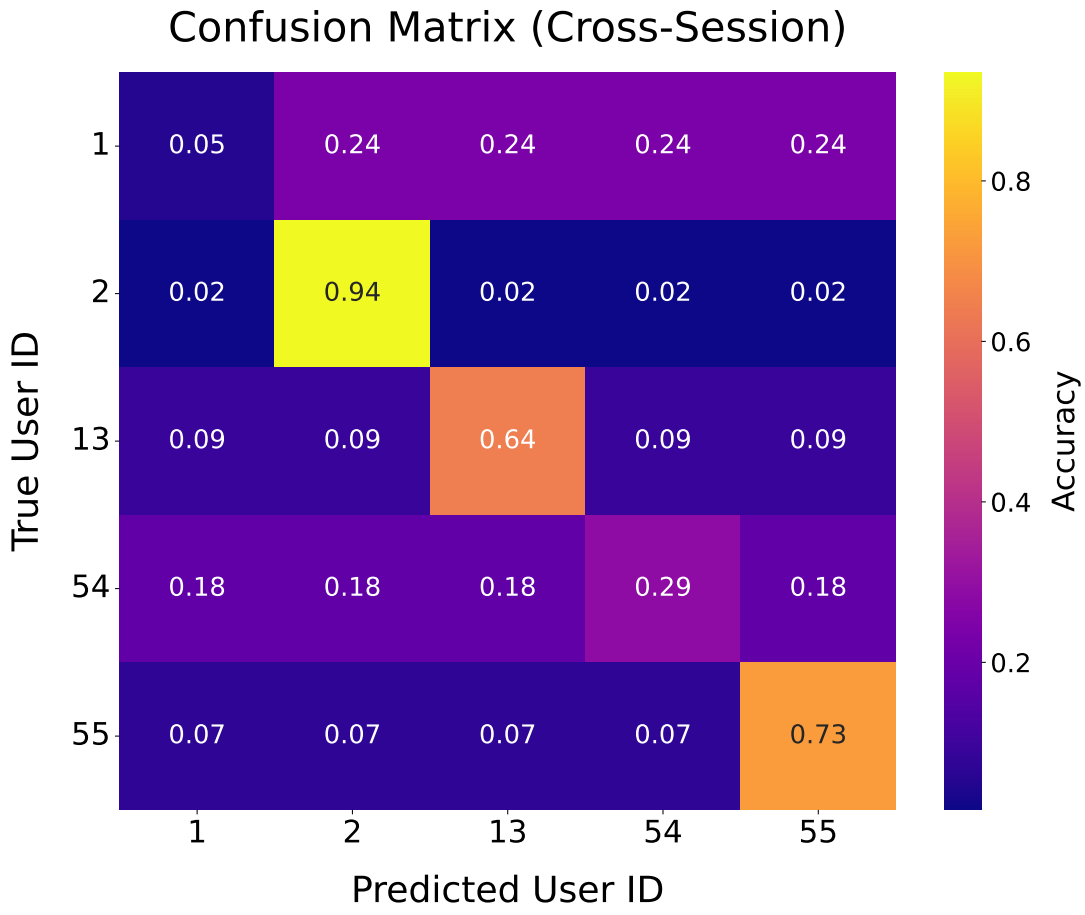


FIGURE 7.6: Normalized confusion matrix for cross-session (unseen) user identification. Diagonal entries indicate correct classification rates per user; off-diagonal values reflect misclassification rates.

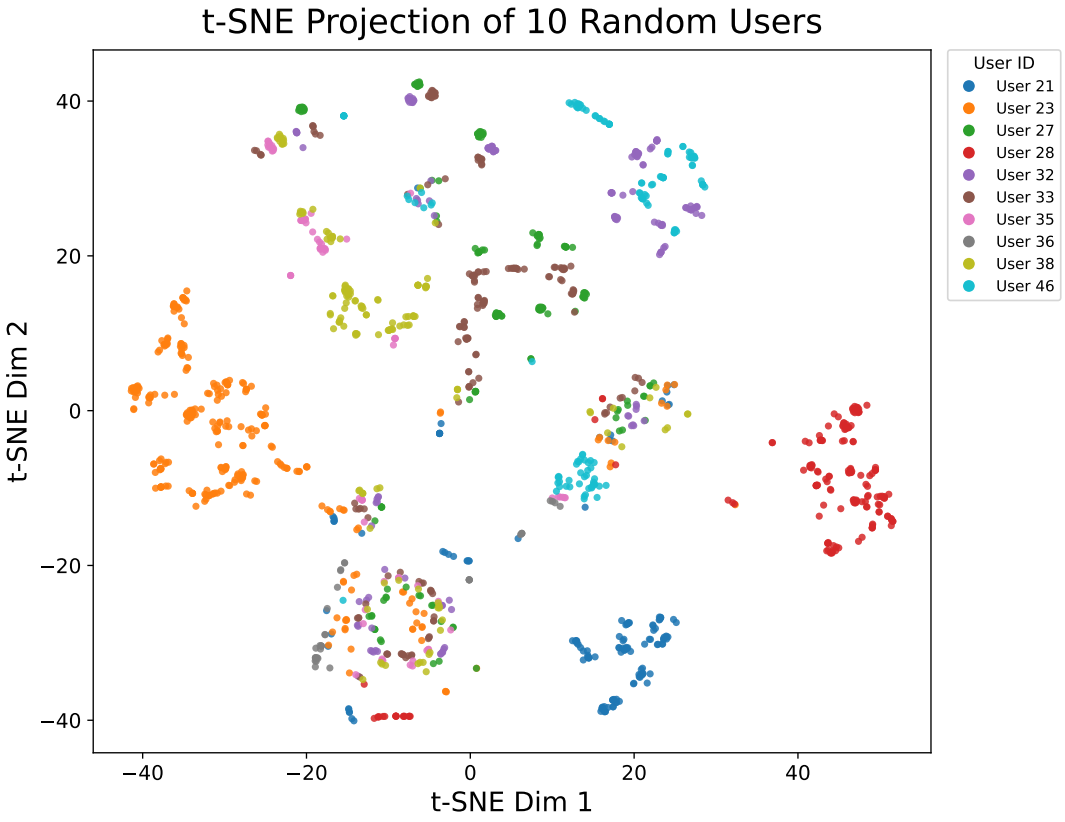


FIGURE 7.7: t-SNE projection of the learned multimodal embeddings for ten randomly selected users. Each point represents a fused gaze, head, and hand sequence; colour denotes the user ID. Distinct, compact clusters indicate strong user-specific separability in the latent space.

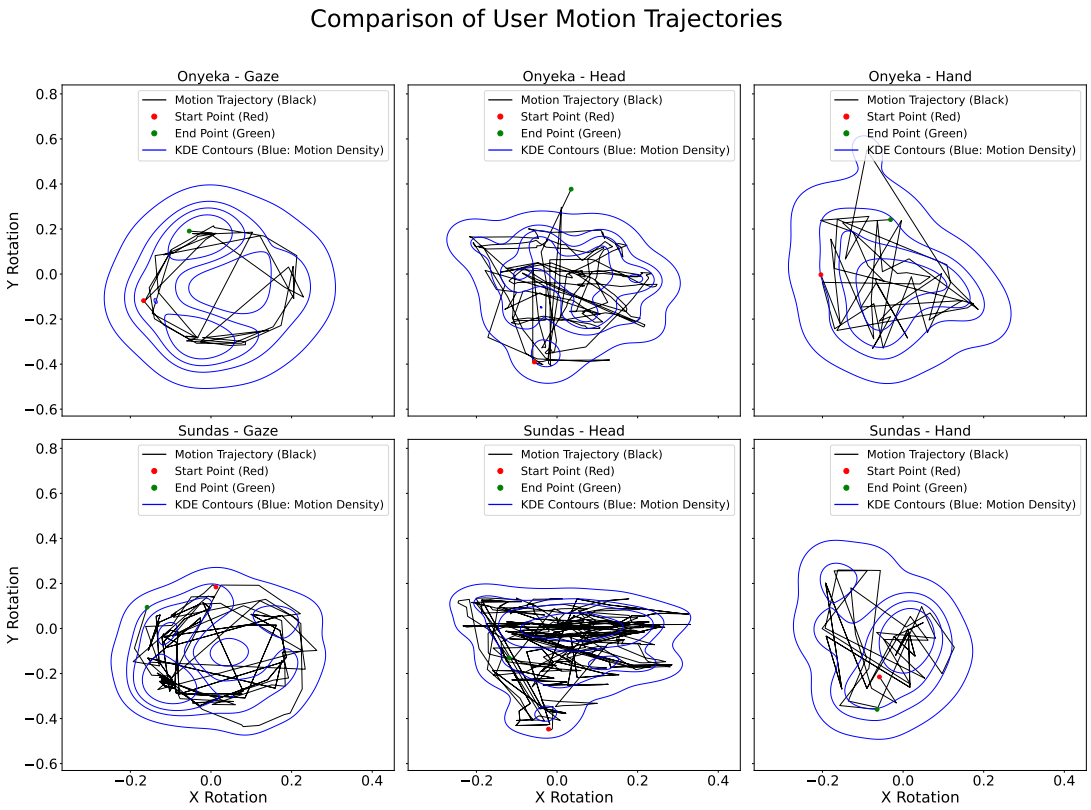


FIGURE 7.8: Comparison of spatiotemporal interaction trajectories for two users across three modalities: eye gaze, head movement, and hand gesture. Each row corresponds to a unique user, showing their behavioral patterns in terms of rotation-based motion (X and Y axes). Black lines indicate movement trajectories, red and green dots mark start and end points respectively, and blue KDE contours reflect motion density. This comparative visualization highlights inter-user differences in movement consistency and spatial dispersion, relevant for user-specific biometric modeling and behavioral variability analysis.

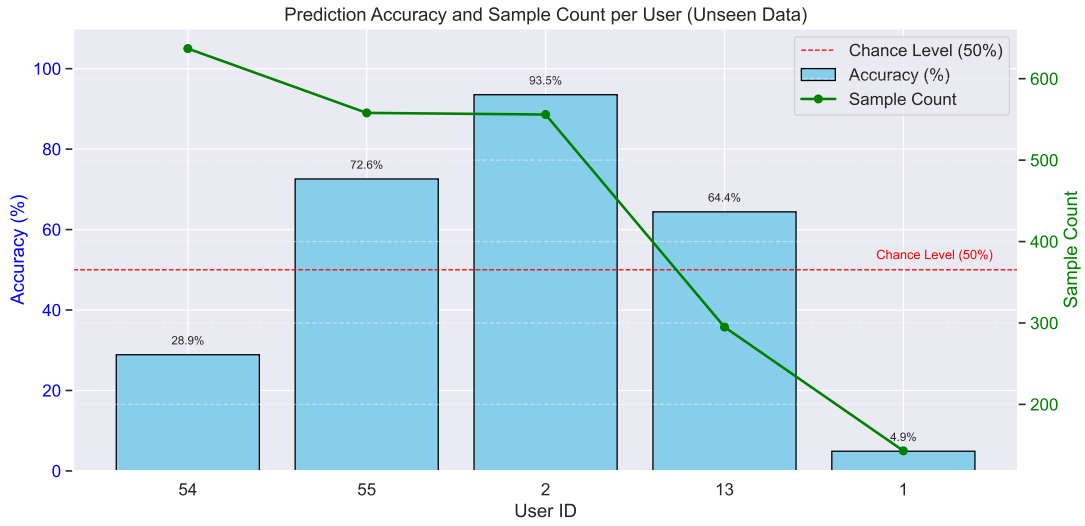


FIGURE 7.9: User-specific classification accuracy and sample count for five unseen users. Accuracy is shown as blue bars, and sample count as a green line. While performance generally correlates with sample availability, individual differences in interaction consistency—such as the contrast between User 13 and User 54—highlight the importance of stable behavioral patterns alongside data volum

identity profiling. However, sample quantity alone does not fully explain the variation in performance. User 13, with significantly fewer samples, was identified with 64.4% accuracy. This discrepancy suggests that familiarity with AR environments and the ability to generate stable and repeatable interaction patterns are crucial to achieving high biometric fidelity, particularly in unseen settings. Users unfamiliar with AR systems may exhibit erratic or inconsistent movement trajectories, thereby diminishing the model's ability to associate behavioural signals with unique identity representations.

These findings underscore two critical insights. First, while larger evaluation sets can aid performance, especially by stabilizing the LSTM-based temporal representations, pattern consistency remains essential. Second, successful generalization in biometric systems may depend not only on the algorithmic architecture but also on human-centric factors, including user proficiency, behavioral variability, and interaction regularity. In practical terms, the most experienced or consistent users tended to maintain higher accuracy when "unseen," whereas users with irregular behaviors saw sharp declines.

To provide further context, Figure 7.10 directly compares the accuracy of the user identification between the trained (intrasession) and unseen (cross-section) scenarios. Although users in the trained setting consistently achieved near-perfect accuracy (approaching 100%), the same individuals exhibited substantial performance declines when evaluated under the unseen condition. This disparity highlights the inherent challenge of generalizing behavioral representations to users not encountered during model development, underscoring the limitations of fixed biometric models in dynamic real-world settings.

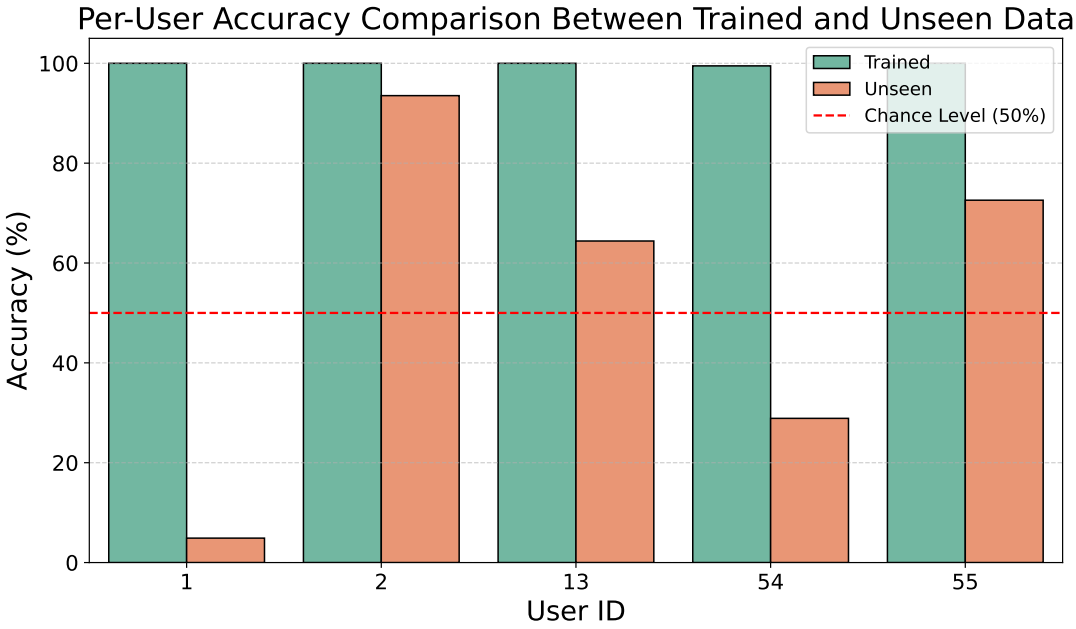


FIGURE 7.10: Comparison of per-user identification accuracy between trained users (intrasession) and unseen users (crossession). Despite perfect recognition during validation, generalization accuracy dropped significantly for certain users, indicating challenges related to behavioral variability, familiarity with AR interaction, and data scarcity.

7.4.2 Performance Metrics Analysis: Accuracy, Precision, Recall, F1 Score, and ROC Curves

To rigorously quantify the framework's effectiveness, we evaluated standard classification metrics for both the validation (seen-user) set and the unseen-user test set. On the validation set, comprising only previously seen users, the model achieved near-perfect performance: overall classification accuracy was 99.74%, with macro-averaged precision, recall, and F1 score of 99.59%, 99.73%, and 99.65%, respectively. Balanced accuracy—averaging per-class recall to account for class imbalance—was similarly high at 99.72%. The macro-averaged ROC AUC (Area Under the Receiver Operating Characteristic curve) reached 99.86%, indicating strong discriminability across identity classes. In contrast, performance on the unseen-user set was lower due to inter-session variability. Per-user classification accuracy in this open-set scenario ranged from 28.9% to 93.5%, depending on each user's behavioral consistency. Aggregated metrics for this group dropped to 25.77% precision, 15.70% recall, and 17.76% F1 score. Balanced accuracy remained substantially above chance (1.82%) at 46.65%, demonstrating partial generalization. Interestingly, the ROC AUC for unseen users remained relatively high at 72.52%, suggesting that while the classifier often misassigned classes, its output probabilities still retained useful identity-discriminative information.

These trends are visualized in the confusion matrix (Figure 7.5), which shows minimal off-diagonal activity for the validation set, confirming that misclassifications were rare and typically limited to a small subset of similar users. The macro-averaged

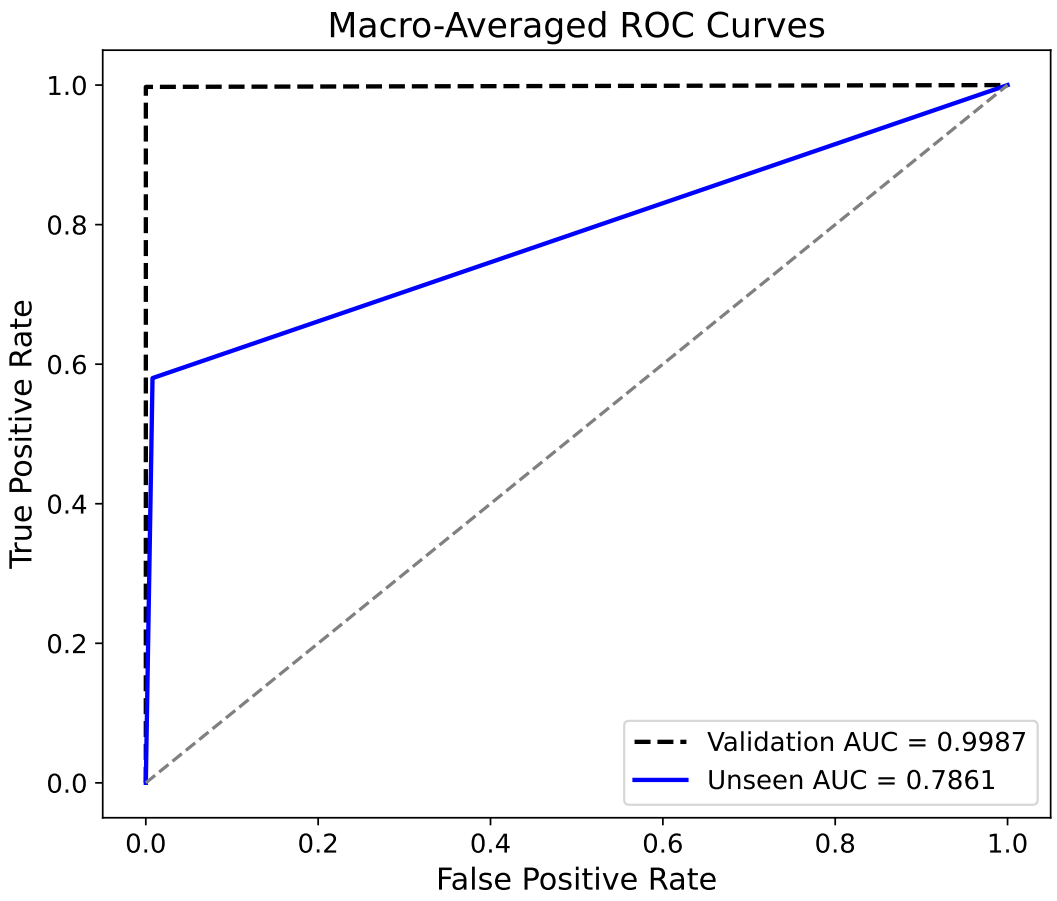


FIGURE 7.11: Macro-averaged ROC curves for validation (dashed black) and unseen users (solid blue). The validation set achieves an AUC of 0.9987, indicating near-perfect class separability for known users. The AUC of 0.7861 reflects moderate class separability when generalizing to identities not observed during training.

ROC curves for both validation and unseen-user sets (Figure 7.11) further illustrate these findings: the validation curve adheres closely to the upper-left ROC boundary, while the unseen-user curve rises more gradually, reflecting diminished separability.

These results highlight the strong performance of the framework under known user conditions and its limited but non-trivial ability to generalize to new users. The observed performance gap underscores the inherent challenge of biometric generalization across sessions, where user behavior may shift due to environmental, cognitive, or device-related factors.

7.4.3 Statistical Evaluation of Validation and Generalization Performance

To assess the consistency and robustness of the proposed multimodal LSTM-based user identification framework, we conducted a per-user statistical analysis of classification accuracy across both the validation and unseen-user sets. For the validation set (known

users), the model exhibited near-perfect consistency, with a mean per-user accuracy of 99.73% (standard deviation 0.99%), and individual users achieving accuracies ranging from 94.74% to 100%. This narrow variance indicates a high degree of reliability in recognising identities for which the model was trained, reflecting the model's effectiveness in capturing stable multimodal behaviour signatures. By contrast, the evaluation on the unseen user dataset revealed greater variability. The mean per-user accuracy dropped to 50.25%, with a standard deviation of 29.15%, and a range from 6.99% to 83.27%. This wider spread illustrates the model's limited generalisation capacity when exposed to novel behavioural patterns not encountered during training. It further suggests sensitivity to individual differences in interaction style and unfamiliarity with the AR interface (as discussed in Section V-A).

To establish the discriminative capacity of the input modalities underpinning our multimodal user identification framework, We performed pairwise Welch's t-tests on the features of ten representative users to quantify how well gaze, head, and hand modalities separate each user pair. The tests confirmed that eye gaze features are consistently among the most discriminative. In many user-pair comparisons, gaze exhibited highly significant differences. For example, users 1 and 3 showed a clear separation in gaze behavior ($t = 7.39$, $p < 10^{-12}$), as did users 1 and 4 ($t = 7.71$, $p < 10^{-13}$) and users 1 and 5 ($t = 6.94$, $p < 10^{-11}$) [268]. In practical terms, user 1 tended to make frequent short saccadic eye movements during interface navigation, whereas users 3 and 5 exhibited longer, more deliberate gaze fixations; this difference was reflected in significantly different variance of their gaze direction time-series. However, several user pairs showed only small or insignificant gaze differences, indicating that gaze alone is not always a unique identifier for every pair of users.

Head movement emerged as a robust biometric signal. Pairwise Welch's t-tests showed consistent differentiation across most user pairs—for example, user 1 vs. user 2 (t = 5.15, p < 10), user 1 vs. user 3 (t = -3.43, $p < 10^{-3}$), and user 8 vs. user 9 (t = 4.62, $p < 10^{-5}$) [269]. Consistent with these statistics, user 8 exhibited higher head-velocity variance and a greater rate of pitch–yaw reversals (markers of rapid nodding and scanning) during interaction, whereas user 9 maintained a steadier head posture.

The hand gesture modality produced more varied results. In some user pairs, hand motion features differed significantly (e.g., user 2 vs. user 3: $t = 4.08$, $p < 10^{-4}$), but in others, no significant difference was found (e.g., user 1 vs. user 3: p = 0.736; user 3 vs. user 4: p = 0.553) [270]. One notable difference was observed between users 6 and 8 ($t = 3.98$, $p < 0.0001$): user 6 favoured smooth lateral hand swipes, while user 8 used shorter, more abrupt pokes or taps – a clear distinction in gesture style that the model could leverage. Overall, however, the hand modality showed the greatest inconsistency in distinguishing users, often being highly context-dependent.

These statistical findings reinforce the role of gaze and head dynamics as the dominant modalities in encoding user-specific behavioural signatures. In contrast, hand gestures – though sometimes informative – appear more context-dependent and less stable as personal identifiers [271]. The consistently strong separation observed for head movement, and the selectively strong signals from gaze, support our multimodal fusion strategy. Figure 7.12 illustrates this via a boxplot of Welch's t-statistics: gaze yields a distribution centered slightly above zero (indicating some positive separability for many pairs), head yields predominantly negative t-statistics (indicating one user

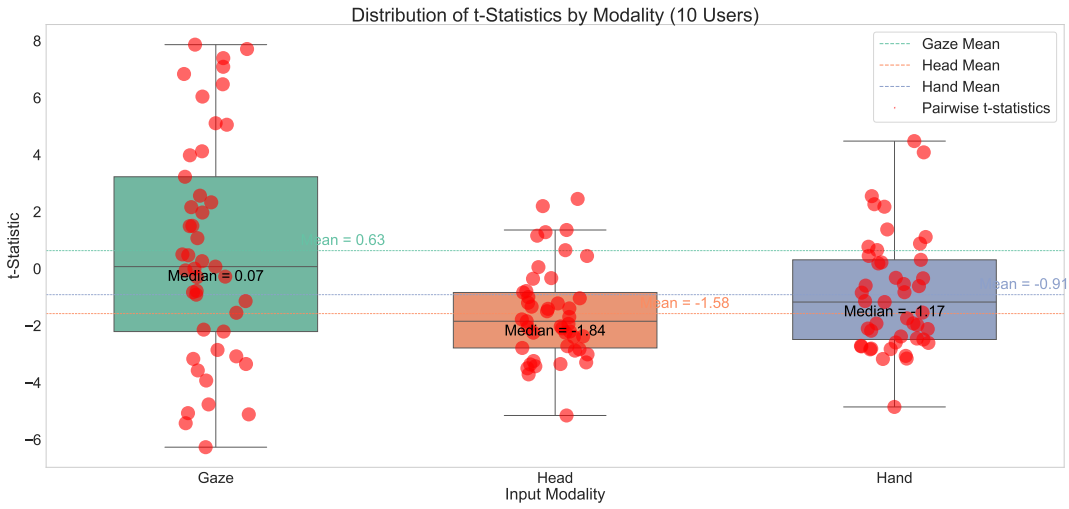


FIGURE 7.12: Boxplot of Welch's t-statistics from pairwise user comparisons across input modalities (Gaze, Head, Hand). Red dots indicate individual t-statistics, with dashed lines showing the mean for each modality. Gaze exhibits greater variance and higher positive t-values in some pairs, while Head and Hand show lower, more negative distributions, reflecting modality-specific discriminative patterns.

often has systematically higher values than another for those features), and hand yields values closer to zero on average with higher variance [272]. No single modality provides a universally sufficient identifier, which validates our decision to combine all three.

7.4.4 Empirical Comparison with State-of-the-Art Modalities

Our analysis confirms that fusing gaze, head, and hand modalities yields near-perfect identification performance for known users and maintains moderate generalization accuracy for unseen subjects. These results underscore the value of multimodal integration for identity-related pattern recognition in augmented reality (AR).

To contextualize our findings, Table 7.2 presents a comparative analysis of recent behavioral biometric systems in AR/VR. Prior studies have typically relied on a single modality and often exhibit performance degradation under cross-session scenarios. For instance, Liebers et al. [243] achieved 90% accuracy using hand-tracking in VR within the same session, but only 38.13% across sessions. Similarly, Asish et al. [252] demonstrated gaze-based user identification using CNN and LSTM models, achieving 85.3% and 84.6% accuracy respectively, though Random Forests outperformed deep models (98.2%) in a single-session setting. Nair et al. [273], using over 55,000 users and 100 seconds of motion data, achieved 94.3% accuracy for closed-set identification in VR. Meanwhile, Jarin et al. [274] reported perfect identification with multimodal models (gaze, hand, body, and face) in same-device conditions, but performance dropped to as low as 5% when tested across devices or applications.

As evident from Table 7.2, our approach outperforms prior unimodal and multimodal methods in unseen-user accuracy. While Jarin et al.'s framework achieves high

TABLE 7.2: Comparison with Prior Behavioral Biometric Methods in AR/VR

Study (Year)	**Platform / Device**	**Modalities**	**Known-user Acc.**	**Unseen-data/across sessions.**
Nair et al. [273] (2023)	VR (Beat Saber gameplay)	Head + Hand motion only	94.3% (100s data)	N/A
Asish et al. [252] (2022)	VR (Vive Pro Eye)	Eye gaze (CNN, LSTM)	98.2% (Random Forest) / 85.3% (CNN)	N/A
Liebers et al. [243] (2021)	VR (Oculus Quest)	Hand-tracking	90% (Session 1)	38.13% (S1→S2)
Jarin et al. [274]	VR (Quest 2, Vive Pro Eye, Pico Neo 3)	Eye + Hand + Body + Face	Up to 100% (same device)	5–40% (cross-device/app)
Our Work	AR (HoloLens 2)	Eye gaze + Head + Hand	**100% (Intrasession)**	**58.0% (Cross-session)**

same-device performance, generalization across hardware and applications remains challenging. In contrast, our method achieves substantially stronger cross-session robustness—58.0% accuracy on unseen users—highlighting the stability and complementarity of fused gaze, head, and hand features in AR.

To better understand the contribution of each modality, we conducted a paired t-test analysis across user samples (figure 7.12). Gaze features consistently produced positive mean t-statistics, confirming their high discriminative power across identity pairs. In contrast, head and hand features yielded predominantly negative or weaker t-values, reflecting lower standalone signal quality. These findings align with prior literature emphasizing the spatial precision and behavioral consistency of gaze-based input [268], [269], [275]. Conversely, head and hand signals are often more susceptible to noise and variability [270].

Recent multimodal systems such as HOIGaze [271] and Wang et al. [272] further support this trend, showing that combining gaze and hand signals improves interaction and recognition in immersive systems. Our findings corroborate this insight: gaze provides the backbone for identity modeling, while head and hand features offer contextual richness that enhances generalization when fused effectively. The accompanying t-test analysis serves as a principled, data-driven justification for this fusion strategy and implicitly confirms that no single modality alone is sufficient to achieve comparable performance.

Our multimodal architecture leverages complementary behavioral signals to achieve state-of-the-art identification performance in AR. The results demonstrate that sensor fusion not only boosts known-user accuracy but also provides clear generalization advantages over prior single-modality and VR-centric methods.

7.5 Conclusion

In this work, we proposed a comprehensive AI-driven behavioural biometric identification framework tailored for augmented reality (AR) environments, with a focus on the Microsoft HoloLens 2 platform. By passively collecting and synchronizing multimodal sensor data, including eye gaze trajectories, head movement kinematics, and hand gesture dynamics, our system constructs rich, time-dependent user signatures that uniquely characterize individuals during natural interaction. These behavioural sequences are processed using a multi-branch Long Short-Term Memory (LSTM) neural network architecture, enabling robust temporal modelling of user-specific patterns across modalities. Unlike prior efforts that relied on explicit tasks, static biometrics, or single modalities, our approach emphasizes implicit, continuous authentication during unstructured AR activity, preserving both usability and security.

Extensive experimental validation across 55 participants demonstrated that our system achieves near-perfect accuracy (over 99%) in intrasession identification and maintains moderate generalizability to unseen users and sessions, with accuracy ranging up to 93.5%. We also presented a thorough statistical analysis confirming the stability and separability of gaze and head movement as dominant biometric features, while highlighting the variability and context dependence of hand gesture data. Visualization through confusion matrices and t-SNE projections further reinforced the system's discriminative power and ability to preserve user-specific identities in the learned feature space. Significantly, our framework operates entirely on-device and requires no transmission of identifiable data, thereby offering a privacy-conscious solution for identity-aware AR systems. We found that a forward pass of the model (which has 2.5 million parameters) on the HoloLens 2 takes under 50 ms for a 3-second window of sensor data, indicating that real-time continuous authentication is feasible on current AR hardware.

Despite the promising performance, challenges remain in generalizing to users not seen during training and across diverse usage scenarios. Behavioral signatures can vary due to factors such as cognitive load, fatigue, familiarity with the interaction, or environmental context, which underscores the need for adaptive and session-invariant modeling strategies. Addressing this variability is crucial for real-world deployment, particularly in collaborative or high-security mixed reality applications where device sharing or identity spoofing risks are significant.

Future Work: As an extension of this research, future efforts will explore decoupling the framework from AR-specific hardware by leveraging conventional RGB cameras or RGB-D sensors to capture behavioural cues in everyday environments. Such a transition would enable user identification systems to operate in more ubiquitous settings—including desktop and mobile interfaces—by extracting behavioural biometrics from video-based observations of head pose, gaze direction, and hand dynamics, without requiring specialized headsets. Additionally, integrating self-supervised learning methods to capture long-term behaviour adaptation, improving resilience to inter-session variability, and extending the system to support open-set identification will be key directions for future development. These advancements would enable scalable, device-agnostic biometric recognition for pervasive computing and ambient intelligence applications.

8 Enhancing Learning in Augmented Reality (AR): A Deep Learning Framework for Predicting Memory Retention in AR Environments

8.1 Introduction

The integration of Artificial Intelligence (AI) and Augmented Reality (AR) has ushered in a new age of human–computer interaction, with far-reaching implications for cognitive science and education. AR overlays digital content onto the physical world, creating immersive environments that transform how learners engage with complex information [37]. However, the impact of AR on memory retention, a cornerstone of durable learning, remains under-explored. Addressing this gap is crucial, because long-term retention ultimately determines the lasting success of any educational or training intervention.

Chen et al., [277] compared MicrosoftHoloLens-based anatomy instruction with traditional slide lectures. Although the AR approach boosted student engagement and reduced test anxiety, it yielded no significant improvement in post-lesson recall. Gargrish etal. [278] reported a similar pattern in geometry: richer visualisation, but limited gains in long-term memory. By contrast, Makhataeva et al. [279] showed that their ExoMem system, which fuses computer vision with AI-driven spatial localisation, cut cognitive load and reduced errors in object-location tasks by more than a factor of seven. Their work, however, focused on spatial memory rather than general knowledge retention.

Memory consolidation is influenced by cognitive, sensory, and environmental factors [280]. Decades of research demonstrate that sustained attention during encoding leads to stronger memory traces [281]. Traditional tools, such as eye-tracking and electroencephalography (EEG), have illuminated portions of this process; however, both face practical limitations in dynamic AR settings. Specifically, eye-tracking cannot infer underlying cognitive states [282], and movement artefacts often contaminate EEG data [283]. Consequently, there is a pressing need for innovative, non-invasive techniques that mesh seamlessly with AR technology.

[0]Publication note: This chapter is based on the peer-reviewed article [276]. https://doi.org/10.1007/978-3-031-97573-8_7.

Beyond formal education, AI-enhanced AR has proven valuable in manufacturing, where live overlays reduce cognitive load and enhance task precision [284]. Most industrial AR pipelines still rely on non-AI tracking and calibration, which limits their adaptability. Meanwhile, recent memory-prediction studies have relied on physiological signals—such as pupil dilation, heart-rate variability, and EEG—yielding promising but invasive solutions [285]. To overcome these constraints, the present research proposes an AI framework that predicts memory retention solely from behavioural interaction data captured by the MicrosoftHoloLens2. By analysing gaze duration, interaction frequency, revisit counts, and head-movement stability, the framework offers a practical and scalable pathway to real-time retention monitoring.

This chapter aligns with the broader aims outlined in Chapter 1 of the dissertation. Earlier chapters (3–6) tackle the spatial side of immersive training—optimising SLAM, reducing latency, and perfecting head-tracking accuracy for AR-based flight simulation. The work introduced here adds the missing cognitive dimension: a non-invasive method for estimating what trainees remember while they are learning. By coupling behavioural memory prediction with high-fidelity AR tracking, the thesis as a whole advances its overarching objective of delivering adaptive, low-load, and measurably effective training environments.

Key contributions of this research

- **AI-driven memory-prediction model** trained exclusively on HoloLens2 behavioural-sensor data, enabling real-time inference of learner retention without biometric instrumentation.
- **Computational framework for non-invasive cognitive modelling** that integrates seamlessly with existing AR tracking and SLAM pipelines.
- **Empirical validation** showing strong agreement between model predictions and self-reported retention, thereby confirming the framework's reliability for adaptive AR learning.

By closing the loop between where a learner looks and what a learner retains, this study lays the groundwork for next-generation AR systems that are not only spatially precise but also cognitively responsive—an essential step toward the dissertation's vision of safer, more effective, and personalised training in aviation and other high-stakes domains.

8.2 Related Works

Memory retention has been extensively examined using eye-tracking and electroencephalography (EEG). Eye-tracking studies show that gaze duration and pupil dilation index attentional focus, a prerequisite for robust memory formation [286], [287]. Kolnes, et al. [288] broadened these findings by demonstrating that pupil dilation mirrors the breadth of attention, reinforcing its value for cognitive-state assessment. Within aviation training, EEG has been linked to brain-wave signatures and reaction time [289] as well as mental fatigue [290]. Across the memory literature, theta and gamma oscillations emerge as principal markers: theta activity across the medial temporal lobe (MTL) and neocortex scales with both accuracy and confidence. In contrast, theta–gamma

phase-amplitude coupling (PAC) binds content-rich memory traces [291]. Expanding on this, Wynnetal. [292] demonstrated that parietal theta power scales with subjective memory confidence, whereas frontal and parietal gamma oscillations predict objective recognition accuracy and decision quality. The authors argue that elevated gamma activity during retrieval supports hippocampal pattern-completion and the reinstatement of neocortical information, providing a mechanistic link between oscillatory dynamics and mnemonic success.

Recent work has transposed these insights into augmented reality (AR). Suzuki, et al. [293] surveyed physiological measures in AR and identified EEG and eye-tracking as the dominant tools for estimating cognitive load. Their meta-analysis concludes that a multi-method protocol—merging EEG, oculometrics, and self-report scales—yields the most reliable estimates, yet remains hampered by motion artefacts and bulky hardware. In parallel, Vortmannetal. [294] trained machine-learning classifiers on synchronized EEG and gaze data to discriminate attention toward physical versus virtual objects, achieving 77% accuracy and illustrating the feasibility of adaptive AR systems driven by real-time neurophysiology. Given the practical obstacles of biosignal acquisition in the wild, several studies now pursue behavioural analytics driven by computer vision and large-scale language-vision models. Shen, et al. [295] introduced a memory-augmentation agent that encodes egocentric video from AR headsets. On the QA-Ego4D benchmark, the system achieved a BLEU score of 8.3, surpassing prior methods (3.4–5.8), and a user study confirmed superior episodic recall relative to baseline human performance. Complementing physiological work, Gargrishetal. [296] compared AR-enhanced geometry lessons with an interactive-simulation (IS) control. Over a two-month retention interval, AR participants maintained substantially higher test scores (post-learning = 12.24; +1 week = 11.76; +2 months = 11.32) than their IS counterparts (9.64, 8.00, 6.44), attributing the advantage to richer embodied interaction and improved visualization of abstract concepts.

Despite these benefits, Suzuki et al. and Vortmann et al. concur that mainstream adoption hinges on lightweight, non-intrusive EEG solutions—technology that remains relatively nascent. To bridge this gap, the present research dispenses with biosignals altogether and exploits the built-in sensors of the Microsoft HoloLens. We extract unobtrusive behavioural metrics—gaze-point dwell time, interaction frequency, revisit counts, and head-movement stability metrics. By training deep-learning models on these contact-free signals, we seek to deliver a practical alternative to biosignal-heavy pipelines and to extend the literature on AR-based cognitive assessment into genuine real-world deployments.

8.3 Material And Methodology

This study presents an AI-based framework for predicting memory retention in augmented reality (AR) environments using deep neural networks (DNN) trained on behavioural interaction metrics. The proposed model is designed to process behavioural data, including gaze duration, interaction frequency, revisit count, and head movement stability, which is captured using Microsoft HoloLens 2 sensors. By leveraging deep learning, this model learns patterns in user interaction behavior and predicts the likelihood of remembering specific objects.

8.3.1 Experimental Setup

The study was designed to develop and validate an AI-based model for predicting memory retention in augmented reality (AR) environments using behavioural data captured through Microsoft HoloLens 2 sensors. The AR environment was developed using Unity 3D and integrated with the Microsoft HoloLens 2 platform. The environment consists of interactive learning tasks designed to simulate real-world educational scenarios, as illustrated in Figure 8.1.

Participants: Thirty-six participants with varying augmented reality knowledge were recruited for the study. Before the experiment, participants received a briefing on the study objectives and provided informed consent.

Procedure: The study was conducted in a controlled AR environment using Microsoft HoloLens 2 to evaluate memory retention through interactive object engagement. Each participant was given the HoloLens 2 headset, which was carefully adjusted to ensure proper fit and calibration. Before beginning the experiment, participants received a brief explanation of the study objectives and were given instructions on interacting with the AR environment. Once the experiment started, various virtual objects were instantiated and positioned within the participant's field of view. Participants were instructed to click on each object to reveal its name, which was displayed in a text field within the AR interface. They were also encouraged to revisit and interact with objects multiple times to reinforce memory retention. This process was designed to simulate real-world learning and recall mechanisms. Throughout the interaction phase, the system continuously recorded the following behavioural data:

Object name: The specific object that the participant interacted with.

Interaction frequency: The number of times a participant clicked on each object.

Revisit count: The number of times a participant returned to a previously interacted object.

Head movement stability: Quantified as the variance of head rotation and position over a fixed time window, used to assess attentional focus and cognitive engagement. The interaction phase lasted for five minutes, during which participants could freely explore the AR environment and interact with the objects as they wished. This ensured that participants had adequate exposure to all objects and could reinforce their memory through repeated interactions.

After the interaction phase, participants received a brief 5-minute pause before completing a free recall form, where they listed objects remembered from the session. No object lists or prompts were provided, ensuring unbiased memory retrieval. No false recalls occurred, though some participants omitted objects, reflecting natural variability. These responses served as ground truth for evaluating the predictions of the AI model.

8.3.2 AI Model Training and Prediction Using Deep Neural Networks (DNN)

This study adopts an efficient approach for predicting memory retention by encoding image identity through category embeddings rather than processing raw image data. The proposed model receives two distinct inputs: behavioural features and object categories. The behavioural features comprise gaze duration, interaction count,

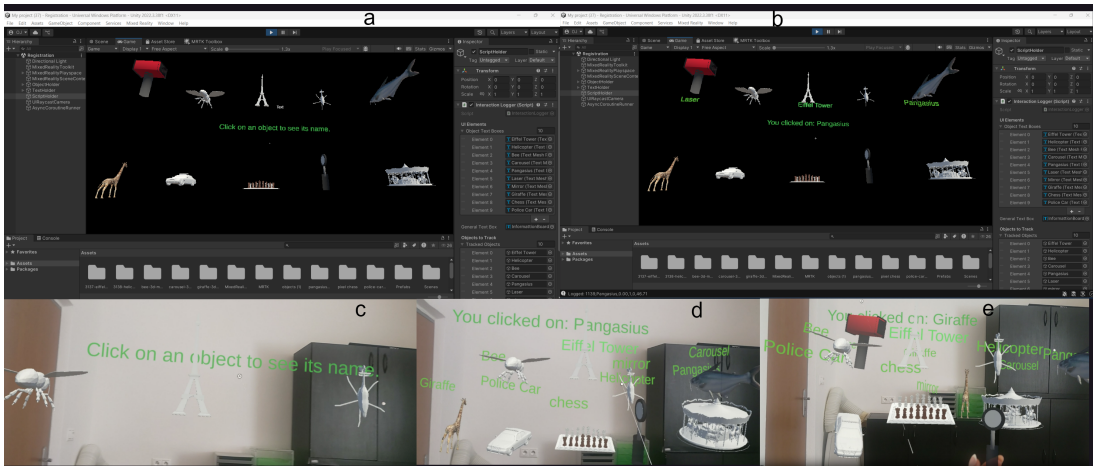


FIGURE 8.1: (a-b) shows the user interface design in Unity for the AR-based memory retention experiment using Microsoft HoloLens 2. (c-e) illustrates the user interactions with objects in the world.

revisit frequency, and head movement, which indicate user engagement within the augmented reality environment. The object category is represented as an integer and mapped into a continuous vector space via an embedding layer. This enables the model to learn dense semantic representations of object classes such as "Bee" or "Carousel."

As illustrated in Figure 8.2, these two input streams are concatenated and processed through a fully connected neural network of two hidden dense layers with ReLU activation functions, followed by a dense softmax output layer. Dropout layers are incorporated to mitigate overfitting and enhance generalization. The final layer of the model is a softmax classifier that outputs one of four memory states: Strong Recall, Weak Recall, Cognitive Overload, or Lack of Engagement. The detailed configuration of the model layers is presented in Table 8.1.

To enhance model performance, Bayesian Optimization is employed to fine-tune key hyperparameters, including learning rate, number of units in each dense layer, batch size, and early stopping patience. The model is trained using the Adam optimizer with categorical cross-entropy as the loss function. Early stopping based on validation loss ensures robust generalization.

This approach offers a lightweight yet semantically rich solution by leveraging category embeddings instead of visual feature extraction. It maintains the capacity to capture object identity while ensuring scalability and practicality for deployment in real-world augmented reality learning environments.

8.4 Results and Discussions

The proposed deep neural network model was evaluated using the accuracy of training, validation, and loss metrics. As shown in Figure 8.3, the model demonstrated stable convergence across training epochs. It achieved a final training and validation accuracy of 0.94 and 0.93, respectively, indicating strong predictive performance on the held-out validation data. The accuracy curves rapidly increased during the initial training

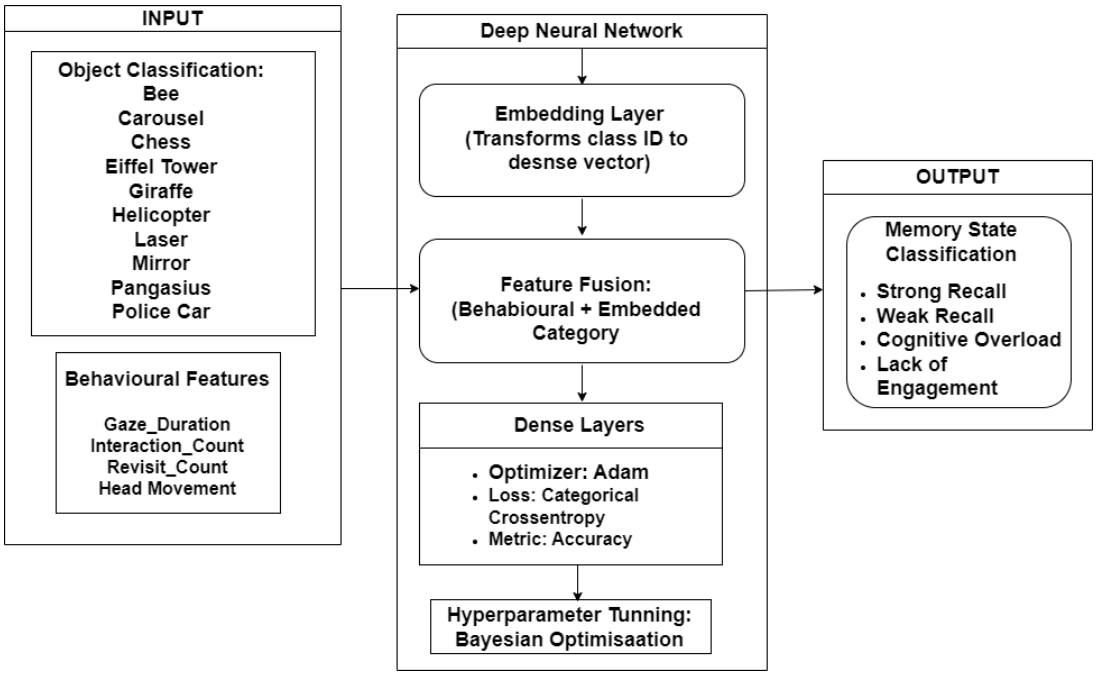


FIGURE 8.2: An overview of the deep learning framework for memory state classification using behavioral features and embedded object categories. Outputs correspond to four memory states, with model parameters optimized via Bayesian optimization.

TABLE 8.1: Model architecture summary.

Layer (type)	Output Shape	Param #	Connected To
input_layer_3 (InputLayer)	(None, 1)	0	-
embedding_1 (Embedding)	(None, 1, 4)	40	input_layer_3[0][0]
input_layer_2 (InputLayer)	(None, 4)	0	-
flatten_1 (Flatten)	(None, 4)	0	embedding_1[0][0]
concatenate_1 (Concatenate)	(None, 8)	0	input_layer_2[0][0], flatten_1[0][0]
dense_3 (Dense)	(None, 20)	180	concatenate_1[0][0]
dropout_2 (Dropout)	(None, 20)	0	dense_3[0][0]
dense_4 (Dense)	(None, 30)	630	dropout_2[0][0]
dropout_3 (Dropout)	(None, 30)	0	dense_4[0][0]
dense_5 (Dense)	(None, 4)	124	dropout_3[0][0]

TABLE 8.2: Categorization of memory states by the AI model based on user interaction metrics.

Predicted Memory State	AI Interpretation (Based on User Interaction Metrics)
Strong Recall	High interaction counts and long gaze duration indicate deep engagement [281]. Frequent revisits reinforce memory encoding, and stable head movement reflects focused attention [287].
Weak Recall	Low interaction counts and short gaze duration suggest brief engagement [287]. Few revisits and stable head movement indicate passive involvement.
Cognitive Overload	Moderate interaction with high revisit counts and frequent gaze shifts implies cognitive strain [288]. Unstable head movement reflects difficulty processing multiple stimuli [293].
Lack of Engagement	Very low interaction, minimal revisits, short gaze duration, and unstable head movement suggest low attention and disengagement [287], [293].

epochs, followed by convergence beyond epoch 80. This trend suggests that the model effectively learned representations from the behavioural interaction features: gaze duration, interaction count, revisit frequency, and head movement alongside the embedded object categories. Correspondingly, both training and validation loss decreased steadily, stabilizing below 0.2, indicating efficient minimization of prediction error. In addition to classification accuracy and loss, the model achieved a mean absolute error (MAE) of 0.14 and a mean squared error (MSE) of 0.404 on the validation set. These low error values indicate that the predicted memory states closely align with the true labels. The results confirm the model's suitability for generalization and effectiveness in memory state classification using embedded object categories and behavioural interaction features. The observed learning behaviour supports the model's potential application in real-time augmented reality (AR)-based learning environments for personalized memory prediction.

Key interaction metrics: Gaze duration, interaction frequency, revisit count, and head movement stability were analysed for their influence on predicted retention scores (Figure 8.4). The results show a strong correlation between engagement and memory retention, with higher scores linked to prolonged gaze, frequent interactions, and revisits. Head movement stability appeared most frequently, likely due to its continuous tracking of subtle attentional shifts. Additionally, gaze duration may influence head stability, as sustained visual focus tends to reduce head movement [287], highlighting the interplay of these features in the model's interpretation of cognitive engagement. A detailed breakdown of the predicted and true memory states across different objects is presented in Figure 8.5. The grouped bar chart reveals that Chess recorded the highest number of predictions and true instances of lack of engagement, followed by Police Car, Pangasius, and Mirror. At the same time, several objects, such as Bee, Laser, and Police Car, also exhibit a considerable number of strong recall cases across both predictions and true states. While this may initially appear

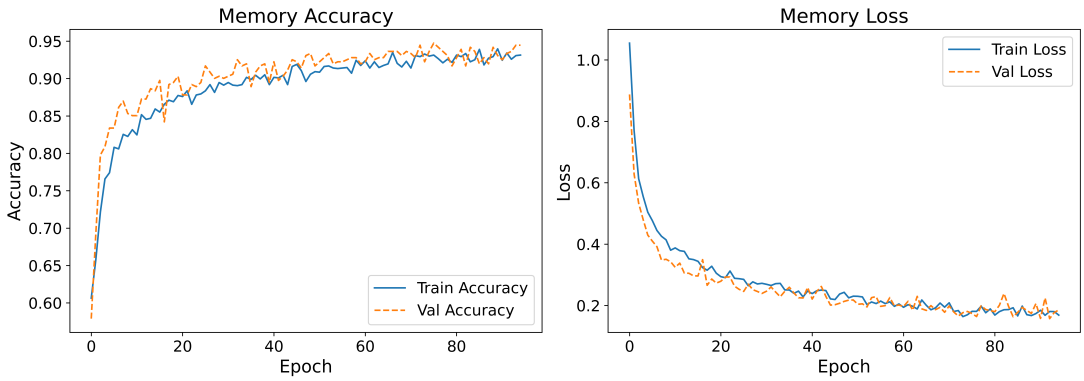


FIGURE 8.3: Training and validation accuracy (left) and loss (right) curves of the deep neural network for memory state classification. The model shows consistent convergence across epochs, indicating effective learning and generalization.

contradictory, it reflects the model's instance-based classification, which evaluates each object interaction independently. These outcomes are informed by distinct behavioural features such as gaze duration, interaction frequency, revisit counts, and head movement stability, captured in real-time through the Microsoft HoloLens 2. Thus, depending on their unique engagement profiles during the interaction, the same object may elicit divergent participant memory state outcomes. For example, users who demonstrated sustained attention and multiple revisits were likely associated with strong recall, while others showing minimal gaze or interaction with the same object often aligned with a lack of engagement. The behavioural variability explains the coexistence of predicted and true memory state discrepancies for a single object. In some cases, memory states observed in the true states were not predicted, and vice versa, highlighting the model's sensitivity to subtle behavioural cues that may not always align with labeled outcomes. Additionally, objects like the Carousel and Police Car showed more frequent weak recall. Meanwhile, cognitive overload remained relatively limited, occasionally appearing in objects such as the Giraffe, Eiffel Tower, and Helicopter. The results reinforce the model's ability to capture and differentiate memory outcomes based on users' heterogeneous behavioural interactions.

8.4.1 Memory State Prediction and User Recall

To evaluate how closely the AI model's memory state predictions reflect actual human memory, we analyzed the predicted memory outcomes for objects that participants recalled after interacting with the AR environment. This comparison provides an empirical basis for assessing the model's cognitive alignment with user recall behaviour.

Figure 8.6 presents a grouped bar chart showing the AI-predicted memory states: Strong Recall, Weak Recall, Cognitive Overload, and Lack of Engagement for objects listed in the user feedback. Overlaid on the chart is a black line indicating the frequency with which users recalled each object. In Figure 8.6, Eiffel Tower was the most frequently recalled object, followed by Pangasius and Laser. The AI model also predicted these objects with a high number of Strong Recall and lower Weak Recall

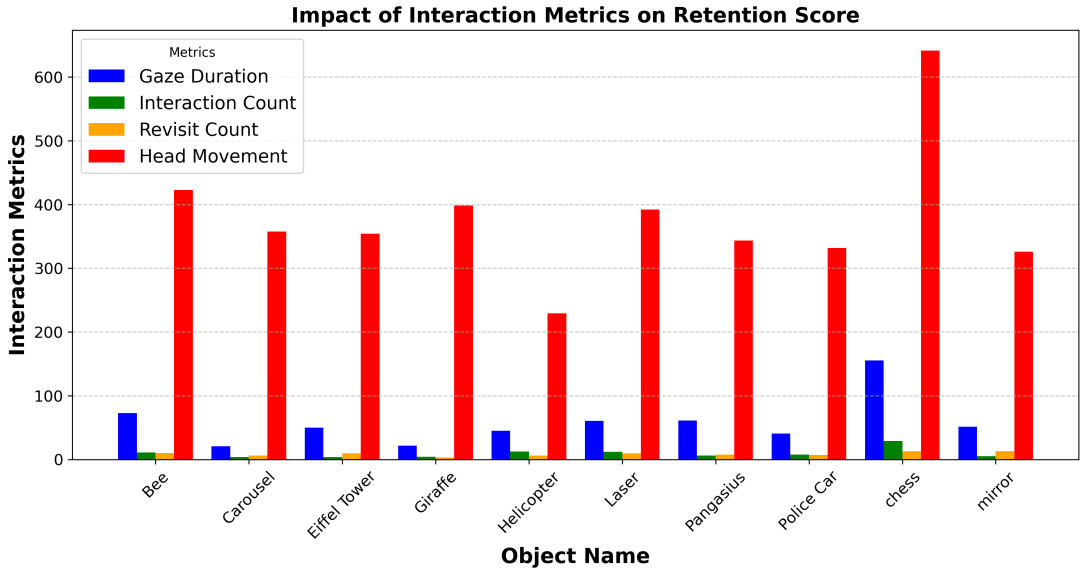


FIGURE 8.4: Impact of interaction metrics: gaze duration, interaction count, revisit count, and head movement on memory retention for different objects.

classifications, demonstrating a precise alignment between predicted memory strength and participant recall. In contrast, Giraffe, Helicopter, and Mirror had the lowest user recall frequencies. The model accurately captured this in the Giraffe case, which received no Strong Recall predictions, only Weak Recall, Lack of Engagement, and Cognitive Overload classifications. This reflects a strong match between the AI's prediction and actual human memory performance, suggesting that the model could infer the object's low memorability based on user interaction features. Although user recall was similarly low for Helicopter and Mirror, the model still predicted some level of Strong Recall. This indicates a partial misalignment and suggests that while the model may have detected surface-level engagement behaviors (e.g., prolonged viewing or revisits), these interactions alone may not always lead to successful long-term recall [287].

Our findings align with research by Mynick et al. [297], which suggests that memory is predictive in perceptual judgments, particularly in immersive environments. Their study found that individuals use memory-based expectations to anticipate visual scenes, enhancing perception and recall efficiency. This supports our observation that users also frequently recalled objects with high AI-predicted retention scores, reinforcing that memory-driven expectations influence engagement and recall patterns.

Similarly, research on spatial memory in augmented reality (AR) environments by Maidenbaum et al. [298] further strengthens this relationship. Their study demonstrated that AR-based spatial memory tasks improve recall accuracy and engagement compared to traditional memory tests, emphasizing the role of interactive and immersive experiences. These findings are consistent with our results, where frequently interacted objects, such as the Eiffel Tower, Laser, and Pangasius, were not only the most accurately recalled by users but also closely aligned with the AI model's predictions.

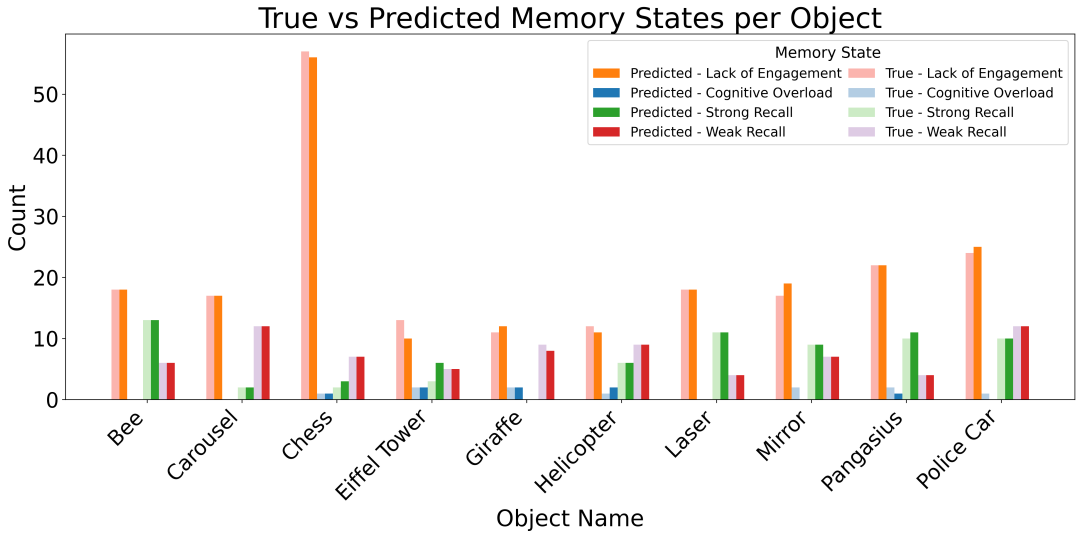


FIGURE 8.5: Grouped bar chart showing AI-predicted and true memory states across objects based on interaction metrics. The distribution reflects object-specific engagement patterns captured during AR experiences.

Furthermore, both studies suggest that familiarity and prior exposure to objects enhance recall accuracy. Mynick et al. [297] highlight that learned environments contribute to faster and more precise recall. This aligns with our observation that AI-predicted memory retention closely mirrored user-reported recall patterns. Maidenbaum et al. [298] extend this idea by demonstrating that spatial engagement and movement enhance memory encoding. This supports our finding that increased interaction frequency and prolonged gaze duration correlate with stronger recall probabilities.

These results validate the AI model's ability to predict memory retention by replicating known cognitive processes. The observed alignment between user feedback and AI predictions suggests that AI-driven memory prediction models can effectively simulate real-world memory retention processes, particularly in AR-based learning and interactive environments.

8.4.2 Statistical Analysis

We conducted repeated-measures ANOVAs with Tukey HSD post hoc tests at the 5% significance level. Normality was confirmed via Q-Q plots (Figure 8.7), validating the ANOVA assumptions by showing that residuals approximated a normal distribution. A one-way ANOVA was conducted to examine differences in interaction features across the ten object categories. The dataset consisted of 1,804 object-level interaction samples (N = 1794).

The results indicated statistically significant differences among objects for all four dependent variables. Interaction Count showed a significant effect, $F(9, N) = 54.48, p < 0.001$. Similarly, significant effects were observed for Revisit Count, $F(9, N) = 15.74, p < 0.001$, Head Movement, $F(9, N) = 36.60, p < 0.001$, and Gaze Duration, $F(9, N) = 65.71, p < 0.001$. Given these significant results, a post-hoc Tukey's

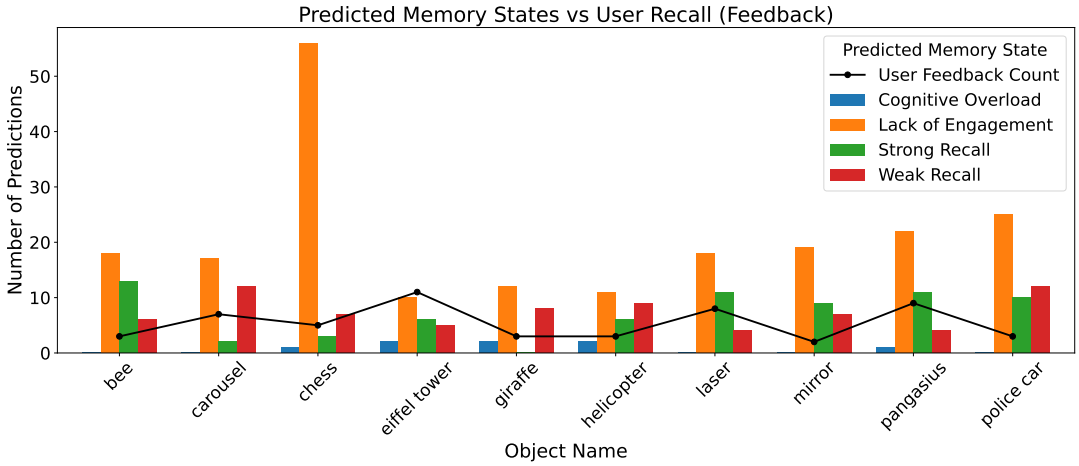


FIGURE 8.6: Comparison of AI-predicted memory states and actual user recall frequencies for recalled objects. The bars represent the predicted memory classification per object, while the black line indicates user recall frequency based on post-interaction feedback.

Honest Significant Difference (HSD) test was performed to determine which objects differed significantly. The analysis included ten distinct objects: Bee, Chess, Pangasius, Mirror, Laser, Eiffel Tower, Helicopter, Giraffe, Carousel, and Police Car.

The Tukey HSD test revealed multiple significant pairwise differences across all dependent variables. Objects with high interaction counts demonstrate significantly different engagement levels, particularly between Chess and Bee ($p < 0.001$). Revisit Count was significantly higher for Mirror and Chess compared to other objects, indicating strong user retention. This finding aligns with the visualization presented in Figure 8.4, which illustrates the impact of interaction metrics on retention scores, highlighting the increased revisit count for Chess and Mirror. Head movement significantly differed between Pangasius and Helicopter, suggesting varying levels of physical engagement. Gaze Duration showed marked differences, with Bee and Chess receiving more visual attention than other objects.

These findings suggest that object type significantly influences user engagement, memory retention, and interaction behaviour. The results show that specific objects elicit distinct cognitive and physical responses, supporting the hypothesis that object features impact user interaction patterns.

8.5 Conclusion

This chapter introduced an AI-driven framework that predicts memory retention in AR by leveraging unobtrusive behavioural signals—gaze duration, interaction frequency, revisit counts, and head-movement stability—captured with the Microsoft HoloLens 2. In doing so, it adds the cognitive layer that complements the thesis's stated focus on spatial fidelity and real-time tracking, thereby advancing the objective of creating adaptive, low-latency, and measurably effective training environments. By confirming (via ANOVA and Tukey HSD) that interaction patterns differ significantly across

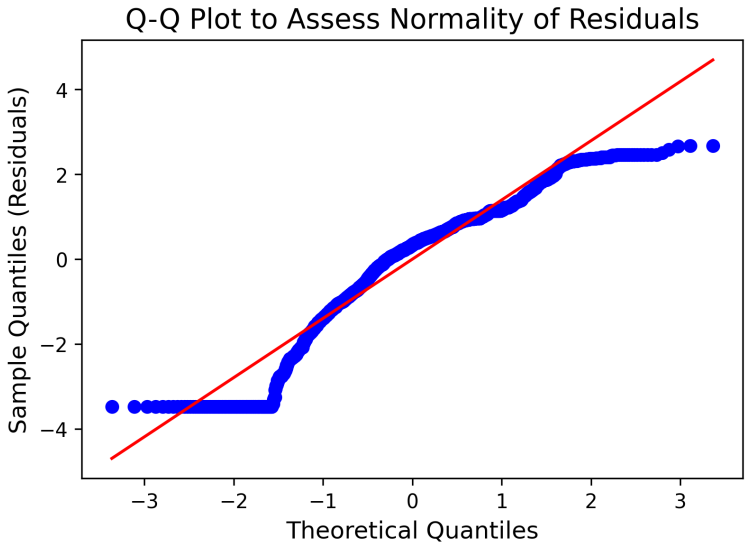


FIGURE 8.7: Q-Q Plot to Assess Normality of Residuals. The points align closely with the theoretical quantiles, indicating approximate normality.

objects, the study empirically links attentional engagement to recall, thereby reinforcing the dissertation's motivation to enhance training realism and efficiency, which can be achieved while reducing pilot cognitive burden in high-stakes simulations.

The framework's contributions—real-time, non-invasive memory prediction from HoloLens behavioural data; seamless integration with existing SLAM/AR pipelines; and strong agreement between predicted and self-reported retention—directly support the dissertation goals of improving AR-based training fidelity and personalization beyond aviation, extending to education, medical training, and workforce development. Looking forward, the following steps align with the dissertation's roadmap: (i) enrich the model with selective physiological or cognitive signals where feasible, (ii) co-optimize memory-prediction and SLAM pipelines for resource-constrained headsets, and (iii) validate the whole system in operational training settings to quantify impact on performance and safety. These directions build on identified SLAM trade-offs and proposed enhancements (e.g., VINS, AKPF fusion) to ensure both spatial and cognitive robustness. This chapter bridges "where the trainee looks" with "what the trainee retains," tightly coupling behavioural analytics to the thesis's spatial tracking advances and thereby fulfilling the stated objectives of delivering immersive, precise, and cognitively responsive AR training systems.

9 Revolutionizing Quantum Learning: Mach-Zehnder Interferometer in Augmented Reality

9.1 Introduction

Quantum computing represents a groundbreaking transformation in computational capability, leveraging the principles of quantum mechanics to address problems that are insurmountable for classical computers [299]. However, the abstract and counterintuitive nature of fundamental quantum concepts—such as superposition [300], entanglement [301], and interference [302]—poses significant challenges for learners. These concepts are foundational for understanding quantum systems; yet, traditional pedagogical methods often rely heavily on complex mathematical formalism, which can hinder comprehension and engagement. Within the broader context of this thesis, which seeks to design an adaptive, SLAM-driven AR framework that reduces cognitive load while maintaining high spatial fidelity, quantum computing serves as an ideal testbed for evaluating whether such a framework can make intrinsically abstract content more accessible. The development of quantum mechanics has been pivotal in advancing our understanding of the microscopic world. From Max Planck's introduction of quantized energy to explain blackbody radiation [303], to Einstein's photon theory of light [304], and Bohr's atomic model [305], the evolution of quantum theory has consistently challenged classical notions of physics. Subsequent milestones, such as de Broglie'sBroglie's wave–particle duality [306], Heisenberg'sHeisenberg's uncertainty principle, and Schrödinger's wave equation, have solidified the framework of quantum mechanics [307]. These theories collectively underpin the behaviour of particles like electrons and photons, yet their abstract nature makes them challenging to convey through conventional educational approaches. Addressing this communication gap aligns directly with the thesis objective of coupling precise spatial tracking with cognitively informed interaction design to enhance learning effectiveness across domains.

Visualization has become a key strategy for enhancing understanding of abstract concepts. In recent decades, optical fibre sensors have gained popularity due to their unique advantages [308], mainly due to intermodal interference [309]. Optical fibre Mach–Zehnder interferometers (MZIs) with femtosecond (fs) laser-fabricated microcavities were introduced in [310]. MZI experiments have long served as practical tools for illustrating quantum interference and wave–particle duality. The Mach–Zehnder

Interferometer (MZI) is a widely used experimental apparatus in quantum optics that helps visualize key quantum phenomena. It splits an incoming photon into two coherent paths using a beam splitter. Mirrors redirect each path and pass through phase shifters before recombining at a second beam splitter. Depending on the relative phase difference, interference patterns emerge, illustrating the principle of quantum superposition, where a photon exists in both paths simultaneously until it is observed. This setup embodies interference, wave–particle duality, and the probabilistic nature of quantum measurement. Various studies have explored its dual-geometric framework, intrinsic and extrinsic, in quantum systems. The study presented by [311] delves into the distinctions between intrinsic geometry, defined by the internal properties of quantum states, and extrinsic geometry, which pertains to how these states are embedded in a larger Hilbert space. Similarly, Avdoshkin and Popov [312] discuss a geometric approach to understanding the quantum states of light within such interferometers, highlighting the role of internal photon trajectories and component alignment in achieving coherent photon propagation and precise control of quantum states.

Despite its educational value, implementing the MZI in physical laboratories is challenging due to factors like photon loss, fabrication imperfections, and the need to preserve coherence. These practical constraints limit accessibility and consistency in instructional contexts. Our augmented reality (AR) platform—developed to fulfil the thesis goals of robust markerless tracking, low latency, and adaptive feedback—addresses these issues by simulating the MZI in a controlled, interactive environment. Photon trajectories remain coherent, losses are eliminated, and component alignments are idealized, allowing users to explore quantum behaviour without the limitations of experimental error. AR fosters conceptual understanding and procedural reasoning by enabling real-time manipulation of quantum components in a spatially accurate setting. Integrating the AR-based MZI into this chapter operationalizes the thesis objectives: (i) demonstrating that high-fidelity, SLAM-enabled AR can maintain precise spatial alignment in complex instructional setups; (ii) leveraging behavioural and cognitive metrics to personalize learning; and (iii) validating transferability from the thesis's original application domain to conceptually dense STEM education. This introduction, therefore, situates quantum computing education not as a tangent but as a critical proving ground for the thesis's central aim—bridging machine perception and human cognition to create practical, scalable AR learning tools.

9.1.1 Challenges in Learning Quantum Mechanics

Traditional quantum-computing courses lean heavily on mathematical formalism, asking students to reason from dense equations and abstract objects (state vectors, unitary transformations, Dirac notation). For many—especially those without strong mathematical preparation—this generates frustration, disengagement, and cognitive overload cite johnston1998student. Hands-on validation is rarely an option; observing genuine quantum behaviour demands sophisticated, expensive equipment that most classrooms cannot provide. Misconceptions only compound the problem. Learners routinely conflate the uncertainty principle with measurement error or picture superposition," entrenching conceptual mistakes that persist well beyond the introductory level [313]. Overcoming these hurdles requires pedagogy that is both intuitive and interactive, precisely the gap that the present thesis aims to fill with a spatially precise,

cognitively aware AR framework. In this study, we therefore deploy augmented reality to simulate the Mach–Zehnder Interferometer (MZI), a canonical experiment in quantum optics. AR allows users to see and manipulate beam-splitting, interference fringes, and path superposition in real-time, anchored to their physical space. This visualization not only bridges theory and (virtual) experiment but also reduces cognitive load – By replacing symbolic abstractions with embodied, spatial interactions, the AR environment operationalizes the thesis objective of lowering mental effort while preserving conceptual rigour. Leveraging SLAM-driven precision – The same robust markerless tracking pipeline developed in earlier thesis chapters maintains virtual optical elements stably registered, ensuring alignment accuracy that is critical for grasping phase relationships. Moreover, the AR simulation eliminates practical barriers inherent to physical MZI setups—such as photon loss, fabrication tolerances, and coherence maintenance—while remaining cost-effective and scalable for classrooms. In doing so, it extends the thesis framework beyond its original aviation-training testbed to a conceptually dense STEM domain, demonstrating the broader applicability of combining high-fidelity spatial tracking with learner-centred cognition. Through this lens, the chapter not only confronts the pedagogical challenges of quantum mechanics but also exemplifies how the thesis's overarching AR platform can democratize access to complex, otherwise intangible phenomena.

9.2 Related Works

Integrating immersive technologies, especially Augmented Reality (AR) and Virtual Reality (VR), into quantum mechanics education offers transformative potential. These tools enable interactive visualizations of complex, abstract quantum concepts, enhancing student comprehension and engagement.

Tarng and Pei [314] demonstrated that VR modules improved motivation and learning outcomes by enabling interactive virtual experiments of key quantum phenomena, such as the cathode-ray tube and photoelectric effect (Figure 9.1). VR reduced cognitive load and bridged theory with observation. Similarly, Li [315] introduced an AR-based interface for simulating a Quantum Turing Machine, allowing real-time interaction with quantum computational models and bridging classical and quantum paradigms.

Suprapto et al. [316] found AR effective in visualizing atomic models through 3D animations in the PicsAR project, significantly enhancing abstract reasoning skills. Schlummer et al. [317] expanded immersive learning with a Mixed-Reality (MR) system for light polarization experiments using Microsoft HoloLens II, providing real-time interaction and improving conceptual understanding of principles like Malus' Law (Figure 9.2).

Weymuth and Reiher [318] developed a virtual quantum chemistry framework, allowing students to engage intuitively with quantum simulations. This immersive approach improved retention and bridged abstract theory with practical understanding.

Despite these advances, Maclean et al. [319] highlighted a lack of conclusive evidence regarding the effectiveness of immersive tools in quantum computing education. Further research is needed to assess their scalability and long-term impact. Nonetheless, immersive environments support situated learning by simulating lab experiences

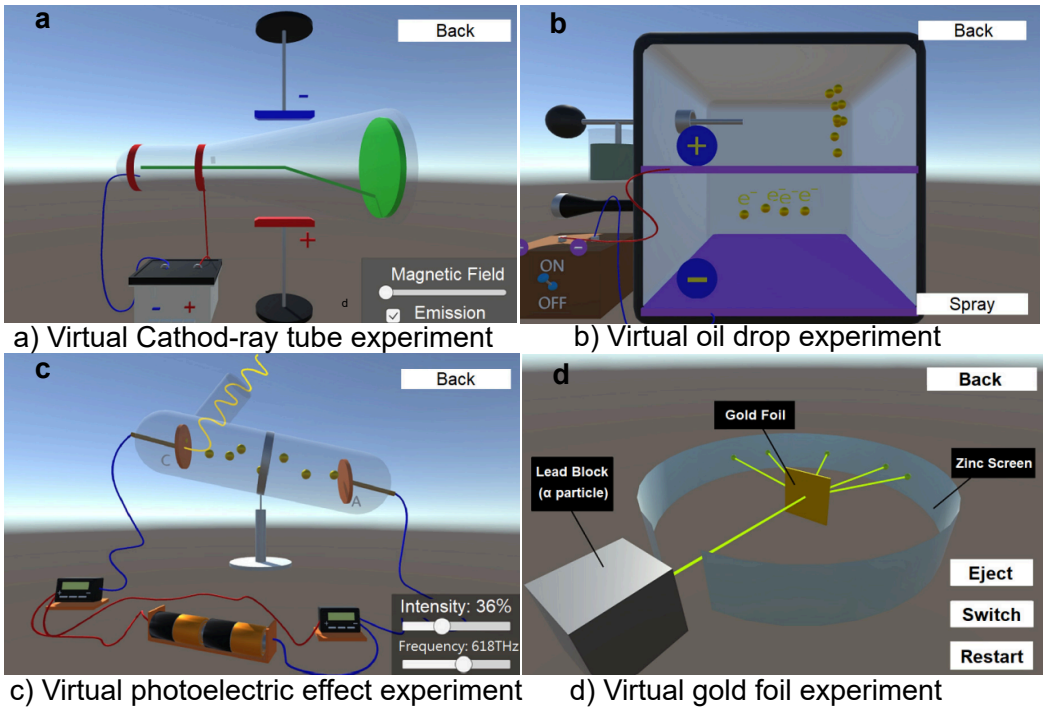


FIGURE 9.1: Virtual simulations of key quantum mechanics experiments designed to enhance interactive learning and visualization. These modules facilitate understanding of fundamental quantum phenomena, making abstract concepts more tangible and engaging [314].

and fostering exploration in a safe, interactive setting. A key gap remains in using AR to visualize core experiments like the Mach-Zehnder interferometer. Addressing this could further enhance quantum education and unlock the full potential of immersive technologies in the classroom.

9.3 Methodology

This study employs a structured experimental methodology to assess the role of augmented reality (AR) in enhancing quantum mechanics education. Central to this investigation is the design and implementation of AR-based visualizations illustrating fundamental quantum optical phenomena, with the Mach-Zehnder Interferometer (MZI) selected for its educational significance. The goal is to address common challenges in quantum mechanics education.

9.3.1 Experimental Setup

This experiment involved developing and deploying an AR-driven platform to showcase the Mach-Zehnder Interferometer, a key quantum optics experiment demonstrating quantum interference and superposition. The study evaluated participants'

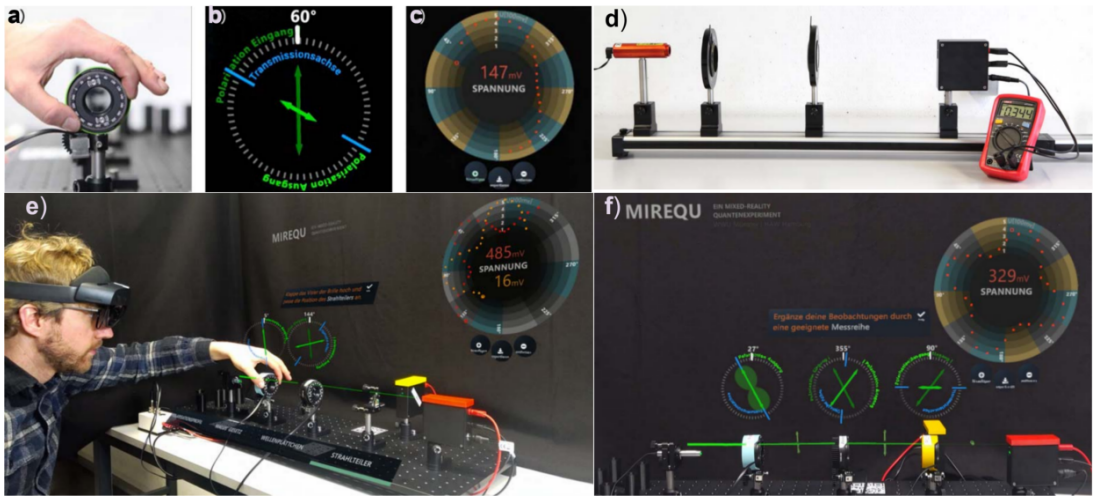


FIGURE 9.2: Mixed-Reality (MR) polarization experiment by Schlummer et al. [317]. (a) Rotation angles of real components as input for visualizations. (b) Polarization vector projection onto a linear polarizer. (c) Light intensity representation in a polar diagram. (d) Traditional polarization experiment setup. (e) Student interaction with MR via Microsoft HoloLens II. (f) Full MR setup visualizing polarization profiles, Malus' Law, and optical effects in real-time.

engagement, comprehension, and retention of these intricate concepts through an immersive AR experience.

Design and Modeling: A 3D model of the MZI was created using Blender and Unity, accurately representing its components, including an optical table, beam splitters, phase shifters, mirrors, and photon detectors, as shown in Figure 9.4.

Marker-Based AR Integration: The AR environment was developed using the Vuforia AR engine, allowing physical markers to trigger the projection of the interferometer onto real-world surfaces. This interactive system enabled real-time manipulation and reinforced visual-spatial understanding.

Procedure Overview: The experimental process was structured into five sequential stages to ensure a consistent learning and assessment experience:

1. **Participant Briefing:** Participants received an overview of the study's objectives and provided informed consent. A short tutorial introduced them to AR interaction basics, ensuring familiarity with the platform's functionality regardless of prior experience.

2. **Instructional Video:** A narrated video demonstrated the operation of the MZI, showing how photons are split and recombined using beam splitters, mirrors, and detectors. This built foundational understanding.

3. **AR Exploration Phase:** Participants interacted with the labelled AR model, allowing them to explore each component's role and observe the resulting photon pathways and interference patterns.

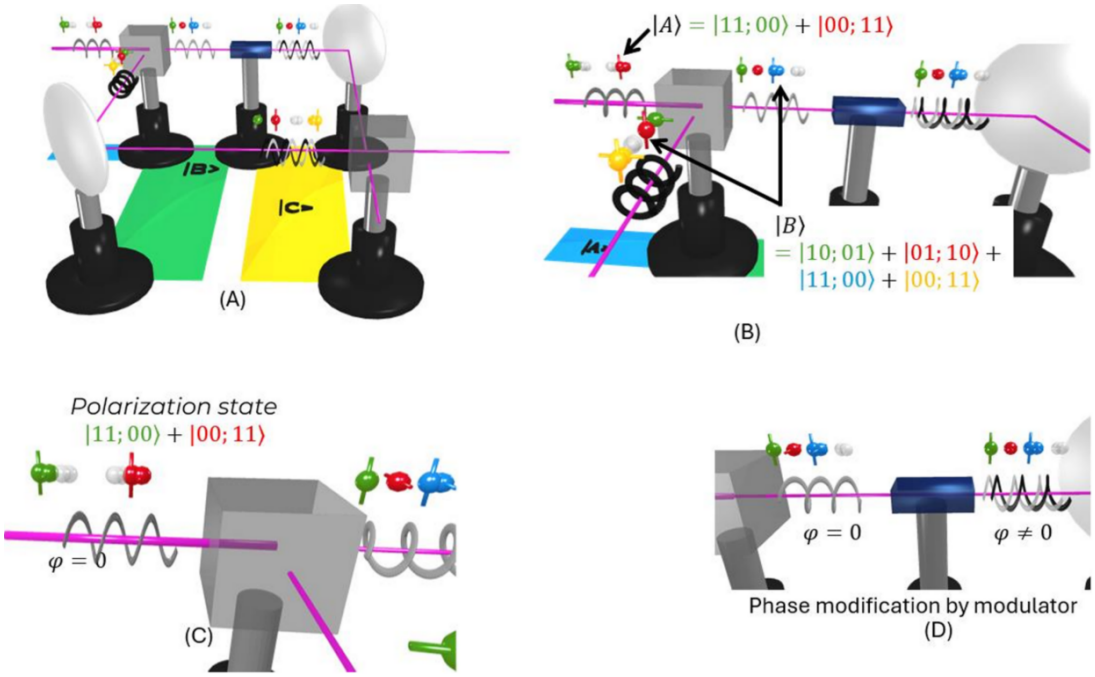


FIGURE 9.3: Visualization of photon trajectories in the Mach-Zehnder Interferometer (MZI), highlighting intrinsic geometries through quantum state evolution and extrinsic geometries via optical component alignment. Photon superposition and phase-modulated pathways are shown in 3D for analyzing quantum interference.

4. **Hands-On Task (Assessment):** Labels were removed from the AR interface. Participants were required to reconstruct the interferometer configuration from memory, testing their conceptual grasp and spatial reasoning.

5. **Post-Interaction Feedback:** Participants rated their comprehension and engagement and provided open-ended feedback. Both quantitative and qualitative data were collected.

Interaction and Evaluation: Twenty participants (6 females, 14 males; aged 16 to 40, $M = 27$, $SD = 8$) with varied prior knowledge of quantum mechanics engaged with the AR platform. While most participants had no prior AR experience, all adapted quickly following the short orientation in Step 1, indicating that the interface presented minimal usability barriers. As detailed above, participants progressed from passive video learning to guided AR exploration to an unguided reconstruction task. This sequencing allowed us to evaluate conceptual understanding and whether the AR system supported procedural recall and spatial reasoning.

Finally, participants provided structured feedback (Step 5), rating the AR platform's intuitiveness, clarity, and educational value using a 5-point Likert scale. These categories were selected to assess both usability and perceived instructional effectiveness. Open-ended reflections were also thematically reviewed to identify patterns related to difficulty, engagement, and cognitive load. Accessibility was supported by removing technical constraints (e.g., no lab equipment, no mathematical formalism),

enabling participants, including those with limited prior exposure to quantum mechanics, to visualize and reconstruct quantum concepts. Notably, even the youngest participants performed the hands-on task with comparable success rates, indicating that the platform's design effectively supports a diverse learner base.

This methodology bridges the gap between theoretical knowledge and practical application, leveraging immersive AR technology to improve accessibility, engagement, and learning outcomes in quantum mechanics education.

Statistical Approach: A paired sample t-test was selected as an appropriate method for comparing pre- and post-interaction scores from the same participants. This analysis approach accounts for the dependent sample design and the continuous nature of the data.

9.4 Results and Discussion

The augmented reality (AR) platform for the Mach-Zehnder Interferometer (MZI) demonstrated significant effectiveness in enhancing quantum mechanics education through immersive interaction. As shown in Figure 9.6, the AR application synchronized physical and virtual environments, allowing participants to interact with components such as beam splitters, mirrors, phase shifters, and photon detectors. This integration provided an intuitive overlay of the virtual quantum optics table onto the physical setup, supporting precise alignment and real-time manipulation. After engaging with the AR simulation, the participants showed a strong understanding of quantum interference and superposition, as evidenced by their ability to reconstruct the interferometer configuration. The platform achieved a 70% success rate in the hands-on task, based on participants scoring four or higher across key post-interaction evaluation metrics.

This success rate was derived from a criterion-referenced evaluation of four core post-interaction metrics: Superposition, Interferance, Wave-particle duality, and the function of the MZI's components. Each was assessed using a 5-point Likert scale from 1 (*Not at all*) to 5 (*Very well*). Participants scoring 4 or above on all four metrics were considered successful, indicating meaningful conceptual and practical understanding.

The success rate was calculated using a symbolic proportion-based formula:

$$\text{Success Rate (\%)} = \left(\frac{n_{\geq 4}}{N}\right) \times 100$$

where $n_{\geq 4}$ represents the number of participants who achieved scores of 4 or higher across all four post-interaction evaluation metrics, and N is the total number of participants. In this study, $n_{\geq 4} = 14$ and $N = 20$, yielding a success rate of 70%. This criterion-based approach reflects best practices in instructional assessment [320], and confirms the AR platform's capacity to bridge theoretical principles with practical application effectively.

User feedback further reinforced the educational impact of the platform. Participants rated their comprehension of quantum concepts before and after the AR experience using a five-level scale from 'Not at all (1)' to 'Very well (5)'. Figure 9.5 displays the distribution of these ratings across categories such as interface intuitiveness,

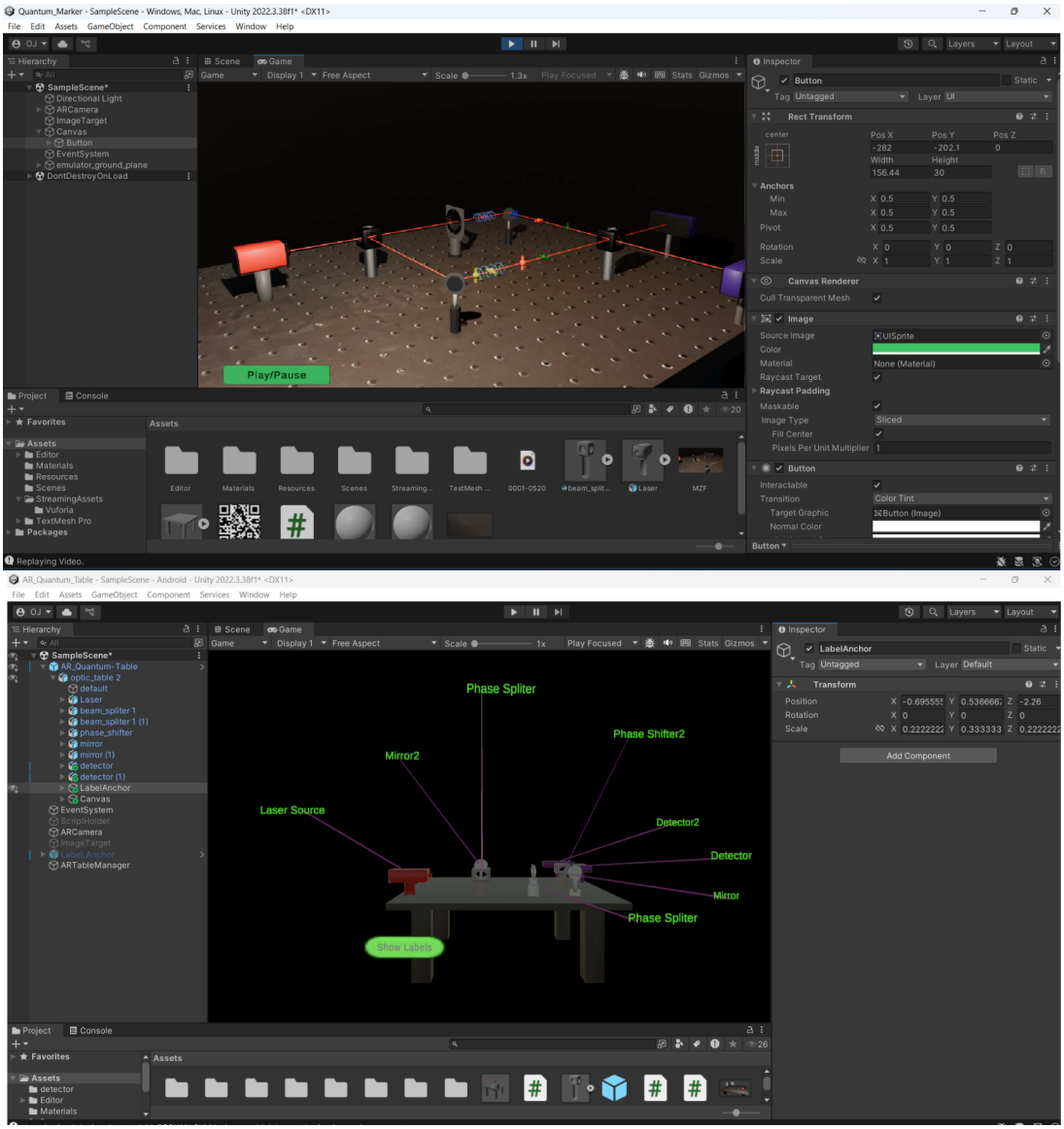


FIGURE 9.4: Unity-based AR application for visualizing the Mach-Zehnder Interferometer. The upper panel shows the design of the instructional video in Unity, while the lower panel displays the quantum table setup with labelled components in the AR application.

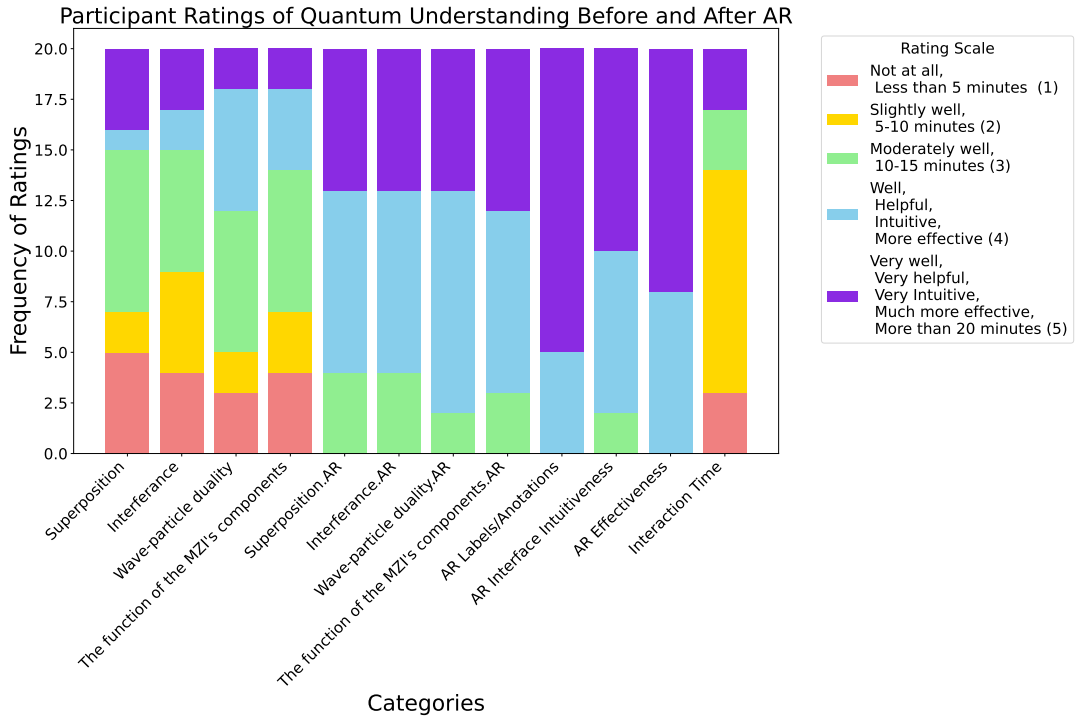


FIGURE 9.5: Participants' ratings of quantum concept understanding before and after AR visualization, categorized by concepts and AR features, ranged from 'Not at all (1)' to 'Very well (5).' Frequencies are shown as stacked bars.

instructional effectiveness, and interaction time. Notably, ratings in the highest category ('Very well, Very helpful, Very Intuitive') increased substantially post-interaction, highlighting improved engagement and understanding.

These findings underscore the transformative potential of AR in quantum education. The platform's dynamic visualization, interactivity, and hands-on reconstruction tasks collectively fostered deeper conceptual engagement, making abstract phenomena more accessible and applicable.

9.4.1 Statistical Analysis

A paired t-test was conducted to evaluate the effectiveness of the augmented reality (AR) platform in enhancing participants' comprehension of quantum concepts. The analysis compared pre- and post-interaction scores across key learning domains, including superposition, wave-particle duality, and the functionality of the Mach-Zehnder Interferometer (MZI) components.

The findings showed a significant statistical enhancement in participants' post-interaction scores ($t(19) = 4.88, p < 0.01$). The mean difference between pre-and post-scores was 1.31 (SD = 1.20), reflecting a measurable enhancement in participants' understanding. On average, pre-interaction scores were 2.89, increasing to 4.20 post-interaction, representing a 45.45% improvement illustrated in Figure 9.7. This

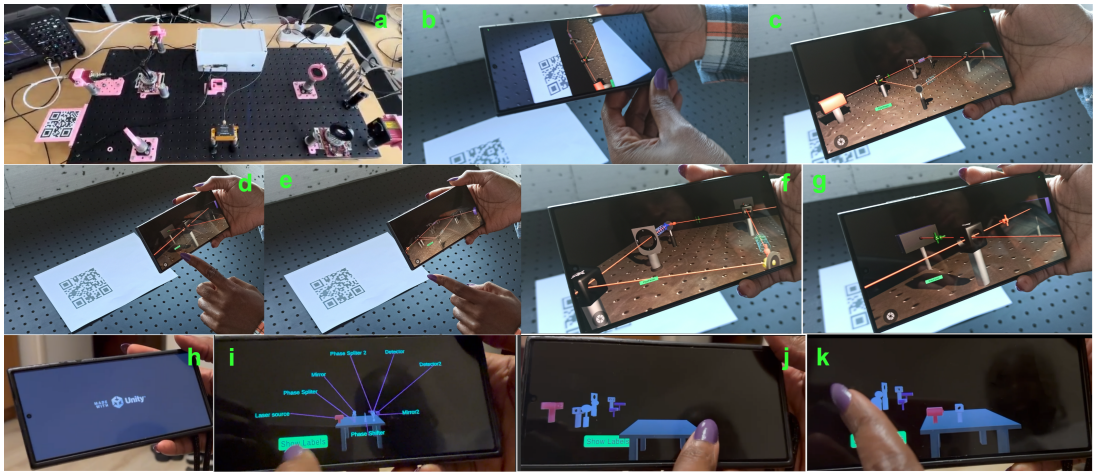


FIGURE 9.6: Augmented reality simulation of the Mach-Zehnder interferometer experiment. (a) Physical setup; (b–f) AR interface displaying interactive photon pathways and component alignment; (g) Unity-based application initialization; (h) AR visualization with labelled components; and (i–k) interactive manipulation and learning tasks.

improvement highlights the platform's capacity to bridge theoretical principles with practical understanding through immersive and interactive learning.

A 95% confidence interval for the mean difference was constructed, ranging from 0.75 to 1.88. This range confirms the reliability of the observed improvement and further supports the robustness of the learning gains facilitated by the AR platform.

These findings underscore the educational effectiveness of the AR platform. The significant increase in comprehension scores, combined with the percentage improvement and narrow confidence interval, highlights its ability to enhance participants' understanding of abstract quantum concepts. This demonstrates the potential of AR technologies to transform STEM education by integrating interactive and immersive learning experiences.

9.4.2 Comparative Analysis

Our study aligns with previous research highlighting the benefits of immersive technologies in science education. [317] demonstrated that mixed reality (MR) significantly enhanced conceptual understanding in polarization experiments, reporting a large effect size ($t(72) = 8.50$, $p < 0.001$, $d = 1.03$) along with high engagement and learner enjoyment. Similarly, Tarng and Pei [314] showed that VR modules enhanced motivation and comprehension in quantum mechanics education.

While these results underscore the potential of MR and VR in improving conceptual outcomes, our AR-based approach offers distinct and complementary pedagogical advantages. Unlike VR and MR platforms that fully immerse learners in synthetic environments, AR uniquely blends digital simulations with the physical world. This allows learners to interact with virtual quantum components overlaid onto real-world

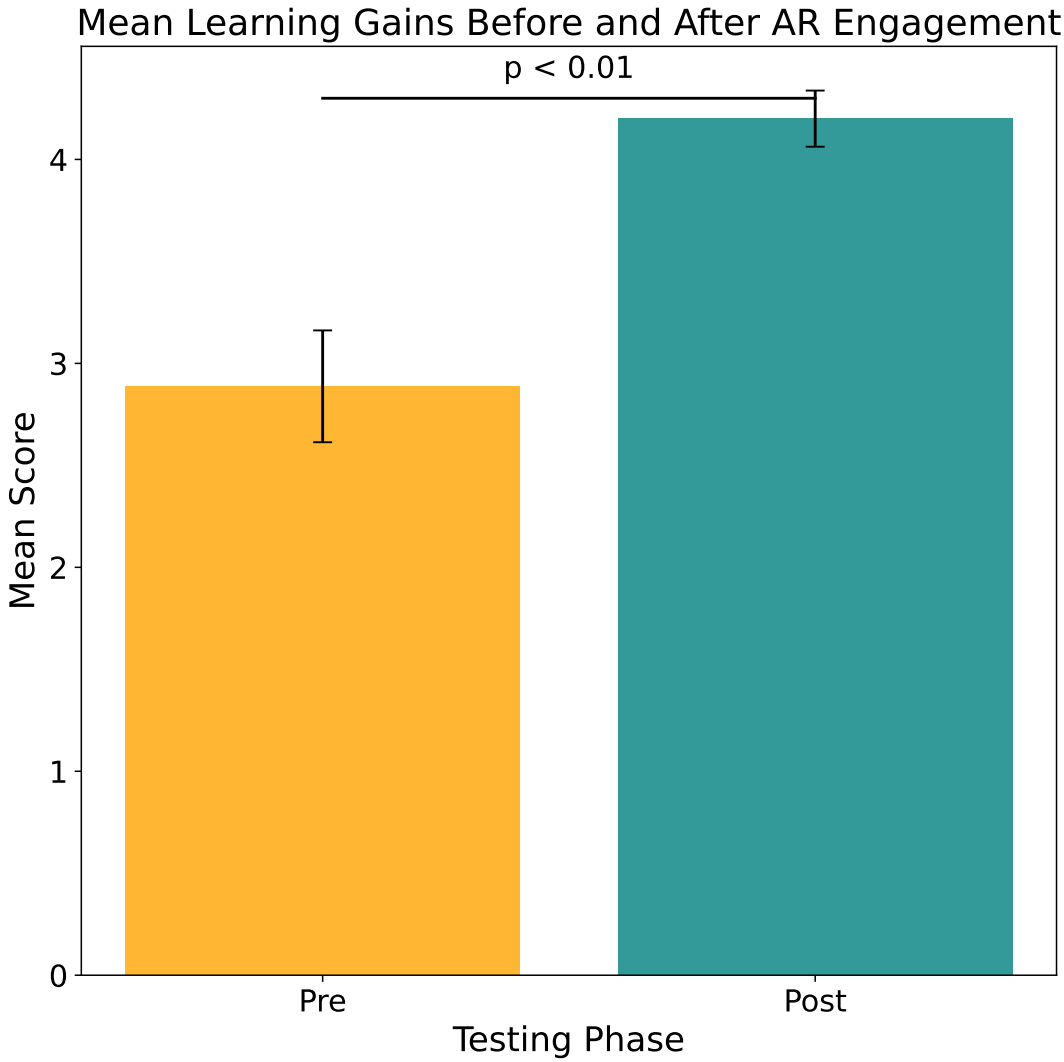


FIGURE 9.7: Mean learning scores before (Pre) and after (Post) AR engagement, showing a significant improvement ($p < 0.01$). Error bars represent the standard error of the mean (SEM).

surfaces, potentially improving spatial contextualization and reducing cognitive dissonance. Whereas VR can isolate users from their physical surroundings, AR maintains real-world reference frames, helping learners relate abstract quantum concepts to tangible spatial representations. Moreover, while many VR platforms focus on scripted, predefined simulations, our AR system enables open-ended exploration and real-time manipulation of Mach-Zehnder Interferometer components. This interactivity was reflected in the 70% hands-on task success rate and the 45.45% gain in post-interaction test scores, suggesting that AR improves both conceptual understanding and procedural competence.

These findings indicate that AR provides a unique bridge between immersive engagement and real-world applications. By fostering embodied learning, supporting spatial reasoning, and enabling authentic task performance, our AR platform contributes a more accessible and context-integrated alternative to fully virtual solutions, particularly for instructional settings involving physical quantum optics experiments.

9.5 Conclusion

Augmented reality (AR) has demonstrated significant potential in enhancing quantum-mechanics education by enabling interactive visualisation of the Mach–Zehnder Interferometer, and the 70% learning gain observed here—together with a statistically significant improvement in participants' understanding of superposition and interference—illustrates how immersive media can transform abstract theory into graspable experience. These outcomes directly serve the thesis aim of reducing cognitive load through embodied, spatially anchored instruction, showing that the same design principles developed initially for high-fidelity, pilot-training scenarios can be transferred to concept-dense STEM content without loss of rigour. The AR environment bridged the gap between theoretical abstraction and practical application, improving concept retention and encouraging sustained engagement; participants repeatedly cited the freedom to manipulate quantum components as instrumental to their conceptual grasp. This aligns with the dissertation's broader vision of a SLAM-driven platform that couples precise overlays with intuitive interaction to democratise access to sophisticated knowledge domains.

While the study confirms AR's pedagogical value, it also highlights the next steps that will further the thesis agenda: assessing long-term retention, scaling deployments across diverse educational contexts, and integrating complementary teaching strategies and richer behavioural analytics to personalise feedback for learners of varying proficiency. Addressing these challenges will not only refine the current prototype but also strengthen the central thesis proposition that a unified, cognitively aware AR framework can deliver durable learning gains across domains where complexity and precision intersect.

10 Conclusion

The culmination of this dissertation underscores the significant advancements achieved in augmented reality (AR) technologies, particularly within flight simulation environments. This body of work focused on refining head-tracking precision by integrating adaptive Kalman-Particle filter fusion methodologies. The comprehensive experiments and analyses conducted throughout this research have successfully validated the efficacy of this approach, establishing new benchmarks for AR performance in dynamically evolving scenarios. In direct support of this vision, Chapter 6 operationalizes the fusion strategy through an Adaptive Kalman–Particle Filter (AKPF) tailored to cockpit dynamics. The AKPF dynamically reweights Kalman and particle estimates based on motion intensity and sensor noise, stabilizing rapid head rotations while preserving responsiveness. This design reduces jitter and effective latency during aggressive manoeuvres, strengthening immersion and enabling longer, more realistic training sessions in AR flight simulation.

The primary objective of this dissertation was to enhance the fidelity of AR-based flight simulation systems by addressing the critical challenges associated with latency, tracking accuracy, and computational efficiency. By developing and applying adaptive Kalman-Particle filter fusion techniques, this research demonstrated a marked improvement in tracking performance, surpassing traditional methods in accuracy and responsiveness. The experimental results substantiated the effectiveness of this fusion approach, showcasing reductions in root mean square error (RMSE) and improved stabilization across position and orientation estimates. A significant contribution of this work is the comparative analysis of diverse tracking methodologies, including Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Ensemble Kalman Filter (EnKF), and particle filter-based methods such as Sequential Importance Resampling (SIR) and Auxiliary Particle Filter (APF). The results highlighted the strengths and limitations of each approach, ultimately demonstrating the superiority of the fused filter system in real-time flight simulation scenarios. This fusion method effectively mitigated the limitations inherent to single-filter models by leveraging the complementary strengths of Kalman and particle filters, resulting in a robust and scalable solution applicable across varied AR environments. Additionally, the experiments yielded low latency and significantly reduced motion sickness, enabling pilots to engage in prolonged practice sessions without experiencing discomfort from latency-induced delays.

Furthermore, the dissertation undertook an extensive review of head-tracking techniques, a critical aspect of AR implementation in flight simulation. The research identified existing challenges in marker-based and inertial measurement unit (IMU)-based systems, including latency, drift, and environmental interference. This review highlights the need for Simultaneous Localization and Mapping (SLAM)-based solutions, which provide real-time, adaptive mapping of the cockpit environment while

accurately tracking pilot head movements. The rigorous evaluation of SLAM methodologies, including ORB-SLAM, RGB-D SLAM, Visual-Inertial SLAM (VINS), and deep learning-enhanced SLAM, formed the backbone of this dissertation's experimental framework.

The dissertation also explored the integration of SLAM methods specifically designed for flight simulation environments. By aligning SLAM technology with the dynamics of cockpit spaces, the research ensured the development of a head-tracking system capable of high-precision tracking even in the presence of reflective surfaces, limited visual references, and rapid user movements. Experimental validations demonstrated that SLAM-based approaches significantly outperformed conventional marker-based systems, yielding improved accuracy, reduced drift, and greater robustness under dynamic conditions. The findings revealed that SLAM's ability to recalibrate in real-time dynamically is vital for maintaining immersive and realistic flight simulation environments, contributing to enhanced pilot training and operational readiness. Another critical dimension of this research involved comparing marker-based and SLAM-based tracking within user training contexts. This evaluation provided empirical evidence highlighting the limitations of marker-based methods, particularly in large and complex environments, such as flight simulators, where occlusions and physical constraints are prevalent. SLAM-based tracking systems, by contrast, demonstrated superior performance in terms of spatial alignment, depth perception, and real-time responsiveness. This dissertation established SLAM tracking as the preferred method for user training, ensuring that virtual overlays remained anchored to physical cockpit elements, facilitating a more seamless and intuitive user experience.

Additionally, adapting Fitts' Law to augmented reality environments provided a theoretical framework for evaluating and optimizing user interactions within immersive 3D spaces. By incorporating head movement and spatial orientation into predictive models, this adaptation effectively reduced interaction task difficulty by approximately 40% across various spatial quadrants. This innovative approach offers valuable insights for designing ergonomic and user-centric AR interfaces, ensuring intuitive and efficient interactions in complex environments.

Complementing spatial fidelity, this study introduces a new technology, a machine-learning identification layer that continuously and implicitly recognizes who is training and what they are doing. A multimodal, multi-branch LSTM fuses time- synchronized eye-gaze, head-motion, and hand-gesture streams into a stable behavioural signature, enabling reliable user and task recognition without explicit calibration. In operational terms, this provides continuous authentication on shared headsets, obviating the need for keyboard- and PIN-based logins, thereby strengthening privacy and access control while preserving the interaction flow. The same identity signal underpins task-contingent personalization: overlays, checklists, and assessments are bound to the correct user and adapted in real-time to the detected activity. Because the method is non-invasive and computationally lightweight for HoloLens-class devices, it is well-suited for safety-critical training, where interruptions and latency are costly. Taken together with the adaptive filtering and SLAM stack, this contribution moves AR flight training from merely precise and stable tracking to systems that are identity-aware and responsive to operational context.

Extending adaptivity beyond identification, this dissertation contributes a

lightweight deep-learning framework that predicts a trainee's memory-retention state from behavioral signals—gaze duration, interaction/revisit counts, and head-movement stability—augmented with category embeddings for the training content. Running on HoloLens 2, the estimator closes the loop between how trainees interact and what they retain by classifying states such as Strong Recall and Cognitive Overload. These inferences support real-time pacing, targeted repetition, and adaptive scaffolding aligned with each user's moment-to-moment cognitive demands during task execution. Because the model is non-invasive and aligns with user-reported recall patterns, it helps maintain comfort and focus. Meanwhile, cognition-aware pacing reduces mental saturation over longer sessions, thereby enhancing training efficiency and strengthening memory consolidation in AR flight simulation.

Extending beyond aviation, the platform's generalisability is illustrated by an AR Mach–Zehnder Interferometer learning module that applies the same spatial, identity-aware, and cognition-adaptive design principles to concept-dense STEM content. The module transforms quantum-optics constructs into SLAM-anchored, manipulable representations of optical elements and beam paths; couples guided exploration with reconstruction to scaffold schema formation; and relies on the same headset-only, privacy-preserving behavioural streams to support user-specific pacing and assistance. This directly advances the dissertation's goal of an AR training stack that is precise, robust, and personalized. This study establishes a reusable template for translating abstract technical knowledge into situated, task-guided interaction, confirming suitability for domains that require both rigorous spatial accuracy and adaptive instruction.

In summary, the findings presented in this dissertation represent a significant stride in advancing AR technologies, particularly in enhancing head-tracking accuracy, interaction modelling, and spatial mapping for flight simulation. The methodologies developed and validated herein offer a robust foundation for future research, inviting further exploration into the scalability and applicability of adaptive filter fusion and SLAM technologies across broader AR domains. As the AR landscape continues to evolve, the contributions of this work are poised to play a pivotal role in shaping the next generation of immersive and responsive AR systems, driving innovation and expanding the horizons of augmented reality applications. This dissertation advances aviation training by addressing critical challenges in flight simulation environments. It paves the way for the widespread adoption of AR in fields that require high-fidelity spatial tracking and real-time adaptability.

Bibliography

[1] J. Rolfe and K. Staples, "The flight simulator as a training device," *Flight Simulation*, pp. 232–249, 1986.

[2] H. Orlady, *Human factors in multi-crew flight operations*. Routledge, 2017.

[3] T. J. Mavin and P. S. Murray, "The development of airline pilot skills through simulated practice," in *Learning through practice: Models, traditions, orientations and approaches*, Springer, 2010, pp. 268–286.

[4] R. T. Hays and M. J. Singer, *Simulation fidelity in training system design: Bridging the gap between reality and training*. Springer Science & Business Media, 2012.

[5] W. M. of Flight, *The 1942 model c-3 link trainer*, Accessed: [Today's Date], 1942. [Online]. Available: https://wmof.com/project/the-1942-model-c-3-link-trainer/.

[6] H. NOWAKOWSKI and J. KOZUBA, "The use of modern simulator technologies in air training on the example of the air force university in dęblin," 2019.

[7] A. Landman, E. L. Groen, M. Van Paassen, A. W. Bronkhorst, and M. Mulder, "Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise," *Human factors*, vol. 59, no. 8, pp. 1161–1172, 2017.

[8] D. Allerton, *Principles of flight simulation*. John Wiley & Sons, 2009, vol. 27.

[9] P. Pradhan, M. Rostami, J. Kamoonpuri, and J. Chung, "The state of augmented reality in aerospace navigation and engineering," 2023.

[10] S. Misra, V. Fraticelli Rivera, N. Khale, and J. L. Albelo Ph D, "Effectiveness of a flight simulation training visual aid for normal and crosswind approach and landing," *International Journal of Aviation, Aeronautics, and Aerospace*, vol. 9, no. 4, p. 5, 2022.

[11] H. Prayitno, M. Cholik, B. Suprianto, I. Qiram, *et al.*, "Impact of flight simulator training on enhancing situational awareness among aviation vocational education cadets," in *2023 International Conference of Computer Science and Information Technology (ICOSNIKOM)*, IEEE, 2023, pp. 1–6.

[12] K. Saastamoinen and K. Maunula, "Usefulness of flight simulator as a part of military pilots training–case study: Grob g 115e," *Procedia Computer Science*, vol. 192, pp. 1670–1676, 2021.

[13] C. Cadena *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[14] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments/MIT press*, 1997.

[15] K. A. Cyran and O. J. Nwobodo, "Mathematical model of hololens 2 goggles kinematics for ar support of flight simulator," *Athens Journal of Technology & Engineering*, vol. 10, no. 3, pp. 175–186, 2023.

[16] O. J. Nwobodo, K. Wereszczyński, and K. Cyran, "Slam methods for augmented reality systems for flight simulators," in *International Conference on Computational Science*, Springer, 2023, pp. 653–667.

[17] D. H. Arjoni, I. de Souza Rehder, J. M. P. Figueira, and E. Villani, "Augmented reality for training formation flights: An analysis of human factors," *Heliyon*, vol. 9, no. 3, 2023.

[18] N. Fezans and T. Jann, "Modeling and simulation for the automation of aerial refueling of military transport aircraft with the probe-and-drogue system," in *AIAA Modeling and Simulation Technologies Conference*, 2017, p. 4008.

[19] L. Lutnyk, K. G. Kim, A. Sarbach, P. Kiefer, R. Häusler, and M. Raubal, "Context-sensitive augmented reality assistance in the cockpit," *International Journal of Human–Computer Interaction*, pp. 1–20, 2025.

[20] M. Maciejewska, P. Kurzawska-Pietrowicz, M. Galant-Gołębiewska, M. Gołębiewski, and R. Jasiński, "Ecological and cost advantage from the implementation of flight simulation training devices for pilot training," *Applied Sciences*, vol. 14, no. 18, p. 8401, 2024.

[21] F. De Crescenzio, M. Fantini, F. Persiani, L. Di Stefano, P. Azzari, and S. Salti, "Augmented reality for aircraft maintenance training and operations support," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 96–101, 2010.

[22] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE computer graphics and applications*, vol. 21, no. 6, pp. 34–47, 2001.

[23] T. Fick, J. van Doormaal, E. Hoving, L. Regli, and T. van Doormaal, "Holographic patient tracking after bed movement for augmented reality neuronavigation using a head-mounted display," *Acta Neurochirurgica*, vol. 163, pp. 879–884, 2021.

[24] F. Vona *et al.*, "Investigating the impact of virtual element misalignment in collaborative augmented reality experiences," in *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, 2024, pp. 293–299.

[25] K. Jambrosic, M. Krhen, M. Horvat, and T. Jagust, "Measurement of imu sensor quality used for head tracking in auralization systems," in *Forum Acusticum*, 2020, pp. 2063–2070.

[26] I. Rabbi and S. Ullah, "A survey on augmented reality challenges and tracking," *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije*, vol. 24, no. 1-2, pp. 29–46, 2013.

[27] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*, IEEE, 2012, pp. 573–580.

[28] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[29] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, IEEE, 2007, pp. 225–234.

[30] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous robot vehicles*, Springer, 1990, pp. 167–193.

[31] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, Springer, 2014, pp. 834–849.

[32] A. I. Mendelsohn, "The united states and the european union in international aviation," *Issues Aviation L. & Pol'y*, p. 13 271, 2004.

[33] C. Jiang, "Recent advances for smart air traffic management: An overview," in *Artificial Intelligence in China: Proceedings of the International Conference on Artificial Intelligence in China*, Springer, 2020, pp. 415–421.

[34] D. Allerton, "The impact of flight simulation in aerospace," *The Aeronautical Journal*, vol. 114, no. 1162, pp. 747–756, 2010.

[35] S. A. Rizvi, U. Rehman, S. Cao, and B. Moncion, "Exploring technology acceptance of flight simulation training devices and augmented reality in general aviation pilot training," *Scientific Reports*, vol. 15, no. 1, p. 2302, 2025.

[36] T. Gangabissoon, G. Bekaroo, and W. Moedeen, "Application of augmented reality in aviation: Improving engagement of cabin crew during emergency procedures training," in *Proceedings of the 2nd International Conference on Intelligent and Innovative Computing Applications*, 2020, pp. 1–8.

[37] O. J. Nwobodo, G. S. Kuaban, T. Kukuczka, K. Wereszczyński, and K. Cyran, "Analysis of marker and slam-based tracking for advanced augmented reality (ar)-based flight simulation," in *International Conference on Computational Science*, Springer, 2024, pp. 208–222.

[38] X. Sheng, S. Mao, Y. Yan, and X. Yang, "Review on slam algorithms for augmented reality," *Displays*, p. 102 806, 2024.

[39] R. Zabels *et al.*, "Reducing motion to photon latency in multi-focal augmented reality near-eye display," in *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II*, SPIE, vol. 11765, 2021, pp. 213–224.

[40] S. C. Mallam, S. Nazir, and S. K. Renganayagalu, "Rethinking maritime education, training, and operations in the digital era: Applications for emerging immersive technologies," *Journal of Marine Science and Engineering*, vol. 7, no. 12, p. 428, 2019.

[41] S. K. Dyer, "Human error and interactions with technology in safety-critical workplaces: Learning from the aviation industry," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2020.

[42] J. Seetohul, M. Shafiee, and K. Sirlantzis, "Augmented reality (ar) for surgical robotic and autonomous systems: State of the art, challenges, and solutions," *Sensors*, vol. 23, no. 13, p. 6202, 2023.

[43] N. Gavish *et al.*, "Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks," *Interactive Learning Environments*, vol. 23, no. 6, pp. 778–798, 2015.

[44] E. Ejichukwu, L. Tong, G. Hazime, and B. Jia, "Enhancing autonomous vehicle design and testing: A comprehensive review of ar and vr integration," *arXiv preprint arXiv:2404.19021*, 2024.

[45] I. Radu, T. Joy, and B. Schneider, "Virtual makerspaces: Merging ar/vr/mr to enable remote collaborations in physical maker activities," in *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, 2021, pp. 1–5.

[46] H.-J. Guo and B. Prabhakaran, "Hololens 2 technical evaluation as mixed reality guide," *arXiv preprint arXiv:2207.09554*, 2022.

[47] B. C. Kress and W. J. Cummings, "Optical architecture of hololens mixed reality headset," in *Digital Optical Technologies 2017*, SPIE, vol. 10335, 2017, pp. 124–133.

[48] T. Feigl, A. Porada, S. Steiner, C. Löffler, C. Mutschler, and M. Philippsen, "Localization limitations of arcore, arkit, and hololens in dynamic large-scale industry environments.," in *VISIGRAPP (1: GRAPP)*, 2020, pp. 307–318.

[49] D. Ungureanu *et al.*, "Hololens 2 research mode as a tool for computer vision research," *arXiv preprint arXiv:2008.11239*, 2020.

[50] S. Aziz and O. Komogortsev, "An assessment of the eye tracking signal quality captured in the hololens 2," in *2022 Symposium on eye tracking research and applications*, 2022, pp. 1–6.

[51] P. Skurowski, D. Myszor, M. Paszkuta, T. Moroń, and K. A. Cyran, "Energy demand in ar applications—a reverse ablation study of the hololens 2 device," *Energies*, vol. 17, no. 3, p. 553, 2024.

[52] S. Teruggi, F. Fassi, *et al.*, "Hololens 2 spatial mapping capabilities in vast monumental heritage environments," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 46, no. 2, pp. 489–496, 2022.

[53] J. C. Navares-Vázquez, Z. Qiu, P. Arias, and J. Balado, "Hololens 2 performance analysis for indoor/outdoor 3fd mapping," *Journal of Building Engineering*, p. 112 826, 2025.

[54] O. J. Nwobodo, K. Wereszczyński, and K. Cyran, "A review on tracking head movement in augmented reality systems," *Procedia Computer Science*, vol. 225, pp. 4344–4353, 2023.

[55] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, "Augmented reality technologies, systems and applications," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 341–377, 2011.

[56] A. G. Berciu, E. H. Dulf, and I. A. Stefan, "Flexible augmented reality-based health solution for medication weight establishment," *Processes*, vol. 10, no. 2, p. 219, 2022.

[57] M. Sırakaya and D. Alsancak Sırakaya, "Augmented reality in stem education: A systematic review," *Interactive Learning Environments*, vol. 30, no. 8, pp. 1556–1569, 2022.

[58] L. Jingen Liang and S. Elliot, "A systematic review of augmented reality tourism research: What is now and what is next?" *Tourism and Hospitality Research*, vol. 21, no. 1, pp. 15–30, 2021.

[59] M. Schumann, C. Fuchs, C. Kollatsch, and P. Klimant, "Evaluation of augmented reality supported approaches for product design and production processes," *Procedia CIRP*, vol. 97, pp. 160–165, 2021.

[60] A. Chouchene, C. A. Ventura, F. Charrua-Santos, and W. Barhoumi, "Augmented reality-based framework supporting visual inspection for automotive industry," *Applied System Innovation*, vol. 5, no. 3, p. 48, 2022.

[61] S. Gül, S. Bosse, D. Podborski, T. Schierl, and C. Hellge, "Kalman filter-based head motion prediction for cloud-based mixed reality," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3632–3641.

[62] Y. Li, B. Ma, H. Chong, and X. Chen, "A benchmark for full rotation head tracking," in *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, 2018, pp. 2106–2111.

[63] H. Li, W. Yan, Y. Yang, Z. Zhao, H. Ding, and G. Wang, "Inside-out accurate head tracking with head-mounted augmented reality device," in *Asian-Pacific Conference on Medical and Biological Engineering*, Springer, 2023, pp. 3–9.

[64] S. Li, K. N. Ngan, R. Paramesran, and L. Sheng, "Real-time head pose tracking with online face template reconstruction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1922–1928, 2015.

[65] G. Papadakis, K. Mania, and E. Koutroulis, "A system to measure, control and minimize end-to-end head tracking latency in immersive simulations," in *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, 2011, pp. 581–584.

[66] G. Lampropoulos, E. Keramopoulos, and K. Diamantaras, "Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: A review," *Visual Informatics*, vol. 4, no. 1, pp. 32–42, 2020.

[67] T. A. Syed *et al.*, "In-depth review of augmented reality: Tracking technologies, development tools, ar displays, collaborative ar, and security concerns," *Sensors*, vol. 23, no. 1, p. 146, 2022.

[68] D. Lee, M. Choi, and J. Lee, "Prediction of head movement in 360-degree videos using attention model," *Sensors*, vol. 21, no. 11, p. 3678, 2021.

[69] J. Liu *et al.*, "Deep learning-based object pose estimation: A comprehensive survey," *arXiv preprint arXiv:2405.07801*, 2024.

[70] R. Ranjan and W.-S. Gan, "Natural listening over headphones in augmented reality using adaptive filtering techniques," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 23, no. 11, pp. 1988–2002, 2015.

[71] R. Chen and S. Guo, "Look-ahead task offloading for multi-user mobile augmented reality in edge-cloud computing," *arXiv preprint arXiv:2305.19558*, 2023.

[72] P. Kumar, S. Chauhan, and L. K. Awasthi, "Human pose estimation using deep learning: Review, methodologies, progress and future research directions," *International Journal of Multimedia Information Retrieval*, vol. 11, no. 4, pp. 489–521, 2022.

[73] S. Jung, Y.-S. Lee, Y. Lee, and K. Lee, "3d reconstruction using 3d registration-based tof-stereo fusion," *Sensors*, vol. 22, no. 21, p. 8369, 2022.

[74] S. Barai and M. Momin, "Outside-in electromagnetic tracking method for augmented and virtual reality 6-degree of freedom head-mounted displays," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2020, pp. 467–476.

[75] M. Singh, R. A. Shankar, and B. Jung, "Inside-out magnetic tracking for virtual/augmented reality applications," *IEEE Sensors Journal*, vol. 21, no. 24, pp. 28 097–28 106, 2021.

[76] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ismar," in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE, 2008, pp. 193–202.

[77] X. Jiang, L. Zhu, J. Liu, and A. Song, "A slam-based 6dof controller with smooth auto-calibration for virtual reality," *The Visual Computer*, vol. 39, no. 9, pp. 3873–3886, 2023.

[78] W. Fang, L. Zheng, and X. Wu, "Multi-sensor based real-time 6-dof pose tracking for wearable augmented reality," *Computers in Industry*, vol. 92, pp. 91–103, 2017.

[79] D. Lee *et al.*, "Vision-based tracking system for augmented reality to localize recurrent laryngeal nerve during robotic thyroid surgery," *Scientific Reports*, vol. 10, no. 1, p. 8437, 2020.

[80] M. Ribo, P. Lang, H. Ganster, M. Brandner, C. Stock, and A. Pinz, "Hybrid tracking for outdoor augmented reality applications," *IEEE Computer Graphics and Applications*, vol. 22, no. 6, pp. 54–63, 2002.

[81] C. Kollatsch and P. Klimant, "Efficient integration process of production data into augmented reality based maintenance of machine tools," *Production Engineering*, vol. 15, pp. 311–319, 2021.

[82] A. Cao, A. Dhanaliwala, J. Shi, T. P. Gade, and B. J. Park, "Image-based marker tracking and registration for intraoperative 3d image-guided interventions using augmented reality," in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, SPIE, vol. 11318, 2020, p. 1 131 802.

[83] J. Rambach, A. Pagani, M. Schneider, O. Artemenko, and D. Stricker, "6dof object tracking based on 3d scans for augmented reality remote live support," *Computers*, vol. 7, no. 1, p. 6, 2018.

[84] A. H. Basori, F. N. Afif, A. S. Almazyad, H. A. S. AbuJabal, A. Rehman, and M. H. Alkawaz, "Fast markerless tracking for augmented reality in planar environment," *3D Research*, vol. 6, pp. 1–11, 2015.

[85] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[86] S. Huang *et al.*, "A new head pose tracking method based on stereo visual slam," *Journal of Visual Communication and Image Representation*, vol. 82, p. 103 402, 2022.

[87] F. Madrigal and F. Lerasle, "Robust head pose estimation based on keyframes for human-machine interaction," *EURASIP Journal on Image and Video Processing*, vol. 2020, pp. 1–19, 2020.

[88] P.-X. Cao, W.-X. Li, and W.-P. Ma, "A tracking registration method for augmented reality based on multi-modal template matching and point clouds," *International Journal of Automation and Computing*, vol. 18, pp. 288–299, 2021.

[89] S.-r. Qu, J. Li, and Y. Shu, "Accurate vehicle location and tracking algorithms based on improved kernelized correlation motion model and kalman filter in intelligent transport surveillance system," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2019.

[90] T. Pintaric and H. Kaufmann, "Affordable infrared-optical pose-tracking for virtual and augmented reality," in *Proceedings of Trends and Issues in Tracking for Virtual Environments Workshop, IEEE VR*, 2007, pp. 44–51.

[91] M. Guri and D. Bykhovsky, "Air-jumper: Covert air-gap exfiltration/infiltration via security cameras & infrared (ir)," *Computers & Security*, vol. 82, pp. 15–29, 2019.

[92] X. T. Trinh, J.-T. Jeng, H.-T. Nguyen, V. S. Luong, and C.-C. Lu, "Two-dimensional position tracking using gradient magnetic fields," *Sensors*, vol. 22, no. 14, p. 5459, 2022.

[93] L. Wöhle and M. Gebhard, "Steadeye-head—improving marg-sensor based head orientation measurements through eye tracking data," *Sensors*, vol. 20, no. 10, p. 2759, 2020.

[94] M. Singh, R. A. Shankar, and B. Jung, "Magnetic position tracking using inductor coils and imu," in *2020 IEEE SENSORS*, vol. 23, 2020, pp. 1–4.

[95] A. Kataria, S. Ghosh, and V. Karar, "Data prediction of electromagnetic head tracking using self healing neural model for head-mounted display," *Science and Technology*, pp. 354–367, 2020.

[96] I. Karaseitanidis and A. Amditis, "A novel acoustic tracking system for virtual reality systems," *Product Engineering: Tools and Methods Based on Virtual Reality*, pp. 99–122, 2008.

[97] C. Dragne, I. Todiriţe, M. Iliescu, and M. Pandelea, "Distance assessment by object detection—for visually impaired assistive mechatronic system," *Applied Sciences*, vol. 12, no. 13, p. 6342, 2022.

[98] L. Ge, Q. Zhang, J. Zhang, and Q. Huang, "Acoustic strength-based motion tracking," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–19, 2020.

[99] A. Wang and S. Gollakota, "Millisonic: Pushing the limits of acoustic motion tracking," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–11.

[100] A. Plinge, S. J. Schlecht, O. Thiergart, T. Robotham, O. Rummukainen, and E. A. Habets, "Six-degrees-of-freedom binaural audio reproduction of first-order ambisonics with distance information," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[101] S. Ionut-Cristian and D. Dan-Marius, "Using inertial sensors to determine head motion—a review," *Journal of Imaging*, vol. 7, no. 12, p. 265, 2021.

[102] S. Muchallil, K. Halnum, and C. Griwodz, "Low-latency head-tracking for augmented reality," in *2021 International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 2021, pp. 192–197.

[103] K. Jambrosic, V. Planinec, M. Horvat, and P. Francek, "Precision of inertial measurement unit sensors in head-tracking systems used for binaural synthesis," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 263, 2021, pp. 2634–2645.

[104] A. Jadid, L. Rudolph, F. Pankratz, K. Wu, P. Wang, and G. Klinker, "Multiimu: Fusing multiple calibrated imus for enhanced mixed reality tracking," 2019.

[105] A. T. Maereg, E. L. Secco, T. F. Agidew, D. Reid, and A. K. Nagar, "A low-cost, wearable opto-inertial 6-dof hand pose tracking system for vr," *Technologies*, vol. 5, no. 3, p. 49, 2017.

[106] S. Yoon, H. jeong Lim, J. H. Kim, H.-S. Lee, Y.-T. Kim, and S. Sull, "Deep 6-dof head motion prediction for latency in lightweight augmented reality glasses," in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, 2022, pp. 1–6.

[107] L. Kästner, V. C. Frasineanu, and J. Lambrecht, "A 3d-deep-learning-based augmented reality calibration method for robotic environments using depth sensor data," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 1135–1141.

[108] Y. Xu, C. Jung, and Y. Chang, "Head pose estimation using deep neural networks and 3d point clouds," *Pattern Recognition*, vol. 121, p. 108 210, 2020.

[109] Y. Chen, "Algorithms for simultaneous localization and mapping," *February*, vol. 3, pp. 1–15, 2013.

[110] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.

[111] Y. Nava, "Visual-lidar slam with loop closure," Ph.D. dissertation, KTH Royal Institute of Technology, 2018.

[112] S. Karam, V. Lehtola, and G. Vosselman, "Integrating a low-cost mems imu into a laser-based slam for indoor mobile mapping," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 149–156, 2019.

[113] J.-E. Deschaud, "Imls-slam: Scan-to-model matching based on 3d data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 2480–2485.

[114] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*, Ieee, 2011, pp. 2564–2571.

[115] M. A. Hoffman, "Microsoft hololens development edition," *Science*, vol. 353, no. 6302, pp. 876–876, 2016.

[116] B. Glocker, J. Shotton, A. Criminisi, and S. Izadi, "Real-time rgb-d camera relocalization via randomized ferns for keyframe encoding," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 5, pp. 571–583, 2014.

[117] T. Sun, Y. Liu, Y. Wang, and Z. Xiao, "An improved monocular visual-inertial navigation system," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 728–11 739, 2020.

[118] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[119] H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao, "Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1974–1982.

[120] R. T. Azuma, "The most important challenge facing augmented reality," *Presence*, vol. 25, no. 3, pp. 234–238, 2016.

[121] P. Skurowski, K. Nurzyńska, M. Pawlyta, and K. A. Cyran, "Performance of qr code detectors near nyquist limits," *Sensors*, vol. 22, no. 19, p. 7230, 2022.

[122] R. Maier and D. Cremers, "Rgb-d vision," in *Encyclopedia of Robotics*, Springer Berlin Heidelberg, 2020, pp. 1–11.

[123] J. Zhang, Y. Yao, and B. Deng, "Fast and robust iterative closest point," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3450–3466, 2021.

[124] X. Gao, R. Wang, N. Demmel, and D. Cremers, "Ldso: Direct sparse odometry with loop closure," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 2198–2204.

[125] P.-C. Su, J. Shen, and S. C. Sen-ching, "A robust rgb-d slam system for 3d environment with planar surfaces," in *2013 IEEE International Conference on Image Processing*, IEEE, 2013, pp. 275–279.

[126] S. Li and D. Lee, "Rgb-d slam in dynamic environments using static point weighting," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2263–2270, 2017.

[127] W. Dai, Y. Zhang, P. Li, Z. Fang, and S. Scherer, "Rgb-d slam in dynamic environments using point correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 373–389, 2020.

[128] L. Cai, Y. Ye, X. Gao, Z. Li, and C. Zhang, "An improved visual slam based on affine transformation for orb feature extraction," *Optik*, vol. 227, p. 165 421, 2021.

[129] B. Bescos, J. M. Fácil, J. Civera, and J. Neira, "Dynaslam: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.

[130] J. Cheng, Y. Sun, and M. Q.-H. Meng, "Improving monocular visual slam in dynamic environments: An optical-flow-based approach," *Advanced Robotics*, vol. 33, no. 12, pp. 576–589, 2019.

[131] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *2014 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2014, pp. 15–22.

[132] P. Bergmann, R. Wang, and D. Cremers, "Online photometric calibration of auto exposure video for realtime visual odometry and slam," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 627–634, 2017.

[133] P. Liu, X. Yuan, C. Zhang, Y. Song, C. Liu, and Z. Li, "Real-time photometric calibrated monocular direct visual slam," *Sensors*, vol. 19, no. 16, p. 3604, 2019.

[134] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[135] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of mavs in unknown environments," in *2012 IEEE international conference on robotics and automation*, IEEE, 2012, pp. 957–964.

[136] H. Yin, S. Li, Y. Tao, J. Guo, and B. Huang, "Dynam-slam: An accurate, robust stereo visual-inertial slam method in dynamic environments," *IEEE Transactions on Robotics*, 2022.

[137] Q. Cheng, S. Zhang, S. Bo, D. Chen, and H. Zhang, "Augmented reality dynamic image recognition technology based on deep learning algorithm," *IEEE Access*, vol. 8, pp. 137 370–137 384, 2020.

[138] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[139] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[140] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[141] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1851–1858.

[142] X. Gao and T. Zhang, "Unsupervised learning to detect loops using deep neural networks for visual slam system," *Autonomous robots*, vol. 41, pp. 1–18, 2017.

[143] M. Geng, S. Shang, B. Ding, H. Wang, and P. Zhang, "Unsupervised learning-based depth estimation-aided visual slam approach," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 543–570, 2020.

[144] F. Li *et al.*, "A mobile robot visual slam system with enhanced semantics segmentation," *IEEE Access*, vol. 8, pp. 25 442–25 458, 2020.

[145] L. Zhang, L. Wei, P. Shen, W. Wei, G. Zhu, and J. Song, "Semantic slam based on object detection and improved octomap," *IEEE Access*, vol. 6, pp. 75 545–75 559, 2018.

[146] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[147] K. Tateno, F. Tombari, I. Laina, and N. Navab, "Cnn-slam: Real-time dense monocular slam with learned depth prediction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6243–6252.

[148] R. Li, S. Wang, Z. Long, and D. Gu, "Undeepvo: Monocular visual odometry through unsupervised deep learning," in *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2018, pp. 7286–7291.

[149] S. Vijayanarasimhan, S. Ricco, C. Schmid, R. Sukthankar, and K. Fragkiadaki, "Sfm-net: Learning of structure and motion from video," *arXiv preprint arXiv:1704.07804*, 2017.

[150] J. C. Dibene and E. Dunn, "Hololens 2 sensor streaming," *arXiv preprint arXiv:2211.02648*, 2022.

[151] S. Chae, S. Yim, and Y. Han, "Flight simulation on tiled displays with distributed computing scheme," *Springer*, pp. 1–6, 2012. DOI: 10.1007/978-4-431-54216-2_1.

[152] C. Brown, J. Hicks, C. H. Rinaudo, and R. Burch, "The use of augmented reality and virtual reality in ergonomic applications for education, aviation, and maintenance," *Ergonomics in Design*, vol. 31, no. 4, pp. 23–31, 2023.

[153] M. Rohs and B. Gfeller, *Using camera-equipped mobile phones for interacting with real-world objects*. na, 2004.

[154] Z. Deineko, N. Kraievska, and V. Lyashenko, "Qr code as an element of educational activity," 2022.

[155] L. Naimark and E. Foxlin, "Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker," in *Proceedings. International Symposium on Mixed and Augmented Reality*, IEEE, 2002, pp. 27–36.

[156] R. Ribeiro *et al.*, "Web ar solution for uav pilot training and usability testing," *Sensors*, vol. 21, no. 4, p. 1456, 2021.

[157] J. W. Wallace, Z. Hu, and D. A. Carroll, "Augmented reality for immersive and tactile flight simulation," *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 12, pp. 6–14, 2020.

[158] O. Araar, I. E. Mokhtari, and M. Bengherabi, "Pdcat: A framework for fast, robust, and occlusion resilient fiducial marker tracking," *Journal of Real-Time Image Processing*, vol. 18, no. 3, pp. 691–702, 2021.

[159] Z. Xing, X. Zhu, and D. Dong, "De-slam: Slam for highly dynamic environment," *Journal of Field Robotics*, vol. 39, pp. 528–542, 2022. DOI: 10.1002/rob.22062.

[160] A. Li, J. Wang, M. Xu, and Z. Chen, "Dp-slam: A visual slam with moving probability towards dynamic environments," *Information Sciences*, vol. 556, pp. 128–142, 2021.

[161] M. Safi and J. Chung, "Augmented reality uses and applications in aerospace and aviation," in *Springer Handbook of Augmented Reality*, Springer, 2023, pp. 473–494.

[162] X. Liu *et al.*, "Large-scale autonomous flight with real-time semantic slam under dense forest canopy," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5512–5519, 2022.

[163] C. Nitschke, "Marker-based tracking with unmanned aerial vehicles," in *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, IEEE, 2014, pp. 1331–1338.

[164] M. Toyoura, H. Aruga, M. Turk, and X. Mao, "Mono-spectrum marker: An ar marker robust to image blur and defocus," *The Visual Computer*, vol. 30, pp. 1035–1044, 2014.

[165] D. Hu, D. DeTone, and T. Malisiewicz, "Deep charuco: Dark charuco marker pose estimation," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2019, pp. 8436–8444.

[166] O. Toole and D. Dolben, *Marker detection and tracking for augmented reality applications*, 2014.

[167] D. Jurado-Rodriguez, R. Muñoz-Salinas, S. Garrido-Jurado, and R. Medina-Carnicer, "Planar fiducial markers: A comparative study," *Virtual Reality*, vol. 27, no. 3, pp. 1733–1749, 2023.

[168] S. Pan and X. Wang, "A survey on perspective-n-point problem," in *2021 40th Chinese Control Conference (CCC)*, IEEE, 2021, pp. 2396–2401.

[169] S. Siltanen, "Theory and applications of marker-based augmented reality," 2012.

[170] S. Vashisht, "Exploration of markers in augmented reality: Methodologies and applications," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, IEEE, 2024, pp. 1082–1086.

[171] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[172] A. R. Sahili *et al.*, "A survey of visual slam methods," *IEEE Access*, 2023.

[173] O. Guclu and A. B. Can, "Fast and effective loop closure detection to improve slam performance," *Journal of Intelligent & Robotic Systems*, vol. 93, pp. 495–517, 2019.

[174] R. Giubilato, W. Stürzl, A. Wedler, and R. Triebel, "Challenges of slam in extremely unstructured environments: The dlr planetary stereo, solid-state lidar, inertial dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8721–8728, 2022.

[175] C. Park, H. Cho, S. Park, S.-U. Jung, and S. Lee, "Strategy for creating ar applications in static and dynamic environments using slam-andmarker detector-based tracking.," *CMES-Computer Modeling in Engineering & Sciences*, vol. 131, no. 1, 2022.

[176] J. C. Cheng, K. Chen, and W. Chen, "Comparison of marker-based ar and markerless ar: A case study on indoor decoration system," in *Lean and Computing in Construction Congress (LC3): Proceedings of the Joint Conference on Computing in Construction (JC3)*, 2017, pp. 483–490.

[177] I. Rabbi, S. Ullah, M. Javed, and K. Zen, "Analysis of artoolkit fiducial markers attributes for robust tracking," in *International Conference of Recent Trends in Information and Communication Technologies (IRICT'14).. University Technology Malaysia*, Citeseer, 2014, pp. 281–290.

[178] O. J. Nwobodo, K. Wereszczynski, G. Kauban, P. Skurowski, and K. A. Cyran, "An adaptation of fitts' law for performance evaluation and optimization of augmented reality (ar) interfaces," *IEEE Access*, 2024.

[179] A. Y. C. Nee and S. K. Ong, *Springer Handbook of Augmented Reality*. Springer Nature, 2023.

[180] A. Uriarte-Portillo, M.-B. Ibáñez, R. Zataraín-Cabada, and M.-L. Barrón-Estrada, "Higher immersive profiles improve learning outcomes in augmented reality learning environments," *Information*, vol. 13, no. 5, p. 218, 2022.

[181] I. S. MacKenzie, "Fitts' law," *The wiley handbook of human computer interaction*, vol. 1, pp. 347–370, 2018.

[182] L. D. Clark, A. B. Bhagat, and S. L. Riggs, "Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology," *International Journal of Human-Computer Studies*, vol. 139, p. 102 413, 2020.

[183] E. Triantafyllidis and Z. Li, "The challenges in modeling human performance in 3d space with fitts' law," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–9.

[184] C. Zhang, X. Li, F. Gao, F. Zhou, and L. Xu, "An experimental research on the directivity of fitts' law in human-computer interaction," in *2015 IEEE International Conference on Progress in Informatics and Computing (PIC)*, IEEE, 2015, pp. 226–229.

[185] P. Chakraborty and S. Yadav, "Applicability of fitts' law to interaction with touchscreen: Review of experimental results," *Theoretical Issues in Ergonomics Science*, vol. 24, no. 5, pp. 532–546, 2023.

[186] L. Sambrooks and B. Wilkinson, "Comparison of gestural, touch, and mouse interaction with fitts' law," in *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, 2013, pp. 119–122.

[187] R. Jiang and Z. Gu, "Current theoretical developments and applications of fitts' law: A literature review," in *Advances in Ergonomics in Design: Proceedings of the AHFE 2019 International Conference on Ergonomics in Design, July 24-28, 2019, Washington DC, USA 10*, Springer, 2020, pp. 753–760.

[188] S. García-Vergara and A. M. Howard, "Three-dimensional fitt's law model used to predict movement time in serious games for rehabilitation," in *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality: 6th International Conference, VAMR 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 6*, Springer, 2014, pp. 287–297.

[189] I. S. MacKenzie and W. Buxton, "Extending fitts' law to two-dimensional tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1992, pp. 219–226.

[190] J. Accot and S. Zhai, "Refining fitts' law models for bivariate pointing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2003, pp. 193–200.

[191] A. Murata and H. Iwase, "Extending fitts' law to a three-dimensional pointing task," *Human Movement Science*, vol. 20, no. 6, pp. 791–805, 2001.

[192] T. Grossman, D. Wigdor, and R. Balakrishnan, "Multi-finger gestural interaction with 3d volumetric displays," in *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, 2004, pp. 61–70.

[193] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli, "An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[194] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen, "A fitts' law study of gaze-hand alignment for selection in 3d user interfaces," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–15.

[195] R. Alazrai, Y. Mowafi, and C. G. Lee, "Anatomical-plane-based representation for human–human interactions analysis," *Pattern Recognition*, vol. 48, no. 8, pp. 2346–2363, 2015.

[196] D. M. Mifsud, A. S. Williams, F. Ortega, and R. J. Teather, "Augmented reality fitts' law input comparison between touchpad, pointing gesture, and raycast," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, 2022, pp. 590–591.

[197] O. J. Nwobodo, G. S. Kuaban, V. Nkemeni, K. Wereszczynski, and K. A. Cyran, "A hybrid adaptive filter for head tracking in augmented reality (ar)-based flight simulators," *IEEE Transactions on Computers*, 2025.

[198] J. Blundell and D. Harris, "Designing augmented reality for future commercial aviation: A user-requirement analysis with commercial aviation pilots," *Virtual Reality*, vol. 27, no. 3, pp. 2167–2181, 2023.

[199] S. Hongli, W. Qingmiao, Y. Weixuan, L. Yuan, C. Yihui, and W. Hongchao, "Application of ar technology in aircraft maintenance manual," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1738, 2021, p. 012 133.

[200] N. Nazri, M. Abdullah, and F. Mustapha, "Implementation of augmented reality in aircraft inspection process," in *E3S Web of Conferences*, EDP Sciences, vol. 477, 2024, p. 00 067.

[201] A. Agrawal and J. Cleland-Huang, "Rescuear: Augmented reality supported collaboration for uav driven emergency response systems," *arXiv preprint arXiv:2110.00180*, 2021.

[202] C.-C. Peng, A.-C. Chang, and Y.-L. Chu, "Application of augmented reality for aviation equipment inspection and maintenance training," in *2022 8th International Conference on Applied System Innovation (ICASI)*, IEEE, 2022, pp. 58–63.

[203] W.-C. Li, J. Zhang, S. Court, P. Kearney, and G. Braithwaite, "The influence of augmented reality interaction design on pilot's perceived workload and situation awareness," *International Journal of Industrial Ergonomics*, vol. 92, p. 103 382, 2022.

[204] H. H. Chan *et al.*, "An integrated augmented reality surgical navigation platform using multi-modality imaging for guidance," *PLoS One*, vol. 16, no. 4, e0250558, 2021.

[205] M. Asghari Oskoei, "Adaptive kalman filter applied to vision based head gesture tracking for playing video games," *Robotics*, vol. 6, no. 4, p. 33, 2017.

[206] L. Elhattab, J. Khairalla, R. Al-Attar, S. Albert, N. Shorim, and E. Eliwa, "Augmented reality applications in the automotive industry," in *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, IEEE, 2023, pp. 357–364.

[207] T. Porcino, D. Trevisan, and E. Clua, "Minimizing cybersickness in head-mounted display systems: Causes and strategies review," in *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, IEEE, 2020, pp. 154–163.

[208] W. He, B. Swift, H. Gardner, M. Xi, and M. Adcock, "Reducing latency in a collaborative augmented reality service," in *Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2019, pp. 1–9.

[209] J. Ahn, H. Choi, J. Hong, and J. Hong, "Tracking accuracy of a stereo camera-based augmented reality navigation system for orthognathic surgery," *Journal of Oral and Maxillofacial Surgery*, vol. 77, no. 5, 1070–e1, 2019.

[210] W. Ritter, G. Kempter, and T. Werner, "User-acceptance of latency in touch interactions," in *Universal Access in Human-Computer Interaction. Access to Interaction: 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II 9*, Springer, 2015, pp. 139–147.

[211] M. Khodarahmi and V. Maihami, "A review on kalman filter models," *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 727–747, 2023.

[212] G. S. Walia and R. Kapoor, "Intelligent video target tracking using an evolutionary particle filter based upon improved cuckoo search," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6315–6326, 2014.

[213] Y. Tian, X. Zhou, X. Wang, Z. Wang, and H. Yao, "Registration and occlusion handling based on the fast icp-orb method for augmented reality systems," *Multimedia Tools and Applications*, vol. 80, pp. 21 041–21 058, 2021.

[214] F. Ababsa, M. Mallem, and D. Roussel, "Comparison between particle filter approach and kalman filter-based technique for head tracking in augmented reality systems," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, IEEE, vol. 1, 2004, pp. 1021–1026.

[215] F.-E. Ababsa and M. Mallem, "Inertial and vision head tracker sensor fusion using a particle filter for augmented reality systems," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No. 04CH37512)*, IEEE, vol. 3, 2004, pp. III–861.

[216] N. Bouaynaya, W. Qu, and D. Schonfeld, "An online motion-based particle filter for head tracking applications," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, IEEE, vol. 2, 2005, pp. ii–225.

[217] A. Asperti and D. Filippini, "Deep learning for head pose estimation: A survey," *SN Computer Science*, vol. 4, no. 4, p. 349, 2023.

[218] K. Huang, C. Wang, and W. Shi, "Accurate and robust rotation-invariant estimation for high-precision outdoor ar geo-registration," *Remote Sensing*, vol. 15, no. 15, p. 3709, 2023.

[219] B. O. Kose, "Beyond the screen: Cross-platform ux for wearables, iot, and emerging technologies," *Navigating Usability and User Experience in a Multi-Platform World*, pp. 1–18, 2025.

[220] R. Wu, X. Wen, L. Yuan, H. Xu, and Y. Liu, "Visual tracking based on deformable transformer and spatiotemporal information," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107 269, 2024.

[221] Y. Kim, H. Bang, *et al.*, "Introduction to kalman filter and its applications," *Introduction and Implementations of the Kalman Filter*, vol. 1, pp. 1–16, 2018.

[222] J. J. LaViola, "A comparison of unscented and extended kalman filtering for estimating quaternion motion," in *Proceedings of the 2003 American Control Conference, 2003.*, IEEE, vol. 3, 2003, pp. 2435–2440.

[223] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[224] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential monte carlo methods," *Sequential Monte Carlo methods in practice*, pp. 3–14, 2001.

[225] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter," *Kalman filtering and neural networks*, pp. 221–280, 2001.

[226] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[227] Q. Liu, Y. Tian, Y. Chai, M. Liu, and L. Sun, "Design of unscented kalman filter based on the adjustments of the number and placements of the sampling points," *ISA transactions*, vol. 108, pp. 188–195, 2021.

[228] G. Evensen, "The ensemble kalman filter: Theoretical formulation and practical implementation," *Ocean dynamics*, vol. 53, pp. 343–367, 2003.

[229] M. Katzfuss, J. R. Stroud, and C. K. Wikle, "Understanding the ensemble kalman filter," *The American Statistician*, vol. 70, no. 4, pp. 350–357, 2016.

[230] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE proceedings F (radar and signal processing)*, IET, vol. 140, 1993, pp. 107–113.

[231] S. J. Mckenna and H. Nait-Charif, "Tracking human motion using auxiliary particle filters and iterated likelihood weighting," *Image and Vision Computing*, vol. 25, no. 6, pp. 852–862, 2007.

[232] X. Xu, E. Mangina, and A. G. Campbell, "Hmd-based virtual and augmented reality in medical education: A systematic review," *Frontiers in Virtual Reality*, vol. 2, p. 692 103, 2021.

[233] M. Abbaszadegan, S. Yaghoubi, and I. S. MacKenzie, "Trackmaze: A comparison of head-tracking, eye-tracking, and tilt as input methods for mobile games," in *Human-Computer Interaction. Interaction Technologies: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part III 20*, Springer, 2018, pp. 393–405.

[234] C. Giordano, "Keep your eyes on the road: Ar huds for automotive," *AEM*, vol. 7, 2021.

[235] Z. Liang *et al.*, "An enhanced adaptive ensemble kalman filter for autonomous underwater vehicle integrated navigation," *Drones*, vol. 8, no. 12, p. 711, 2024.

[236] B. Martin, T. Nguyen, and R. Anderson, "Improving signal processing in ar systems with enhanced noise reduction techniques," *IEEE Transactions on Signal Processing*, vol. 68, pp. 345–358, 2020.

[237] R. Manzano, P. Brown, and J. Lee, "Sequential filtering methods for real-time position tracking in augmented reality," *Sensors*, vol. 22, pp. 567–575, 2023.

[238] F. Luo, Q. Wang, and D. Zhang, "Improved unscented particle filter for head tracking in augmented reality," *IEEE Transactions on Industrial Electronics*, vol. 67, pp. 986–993, 2020.

[239] X. Zhao and Y. Sun, "Robust kalman filtering for ar environments: A comparative study," *International Journal of Robotics and Automation*, vol. 29, pp. 467–475, 2014.

[240] K. Helin, T. Kuula, C. Vizzi, J. Karjalainen, and A. Vovk, "User experience of augmented reality system for astronaut's manual work support," *Frontiers in Robotics and AI*, vol. 5, p. 106, 2018.

[241] S. Stephenson, B. Pal, S. Fan, E. Fernandes, Y. Zhao, and R. Chatterjee, "Sok: Authentication in augmented and virtual reality," in *2022 IEEE symposium on security and privacy (SP)*, IEEE, 2022, pp. 267–284.

[242] S. G. Grandhi, "Evaluating vr/ar security through continuous authentication via eye tracking movements," M.S. thesis, University of Windsor (Canada), 2025.

[243] J. Liebers *et al.*, "Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–11.

[244] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson, "Personal identifiability of user tracking data during observation of 360-degree vr video," *Scientific Reports*, vol. 10, no. 1, p. 17 404, 2020.

[245] F. Boutros, N. Damer, K. Raja, R. Ramachandra, F. Kirchbuchner, and A. Kuijper, "On benchmarking iris recognition within a head-mounted display for ar/vr applications," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2020, pp. 1–10.

[246] I. Sluganovic, M. Serbec, A. Derek, and I. Martinovic, "Holopair: Securing shared augmented reality using microsoft hololens," in *Proceedings of the 33rd annual computer security applications conference*, 2017, pp. 250–261.

[247] S. Wang *et al.*, "Augmented reality as a telemedicine platform for remote procedural training," *Sensors*, vol. 17, no. 10, p. 2294, 2017.

[248] M. Moghaddam, N. C. Wilson, A. S. Modestino, K. Jona, and S. C. Marsella, "Exploring augmented reality for worker assistance versus training," *Advanced Engineering Informatics*, vol. 50, p. 101 410, 2021.

[249] Y. Zhang, C. Slocum, J. Chen, and N. Abu-Ghazaleh, "It's all in your head (set): Side-channel attacks on {ar/vr} systems," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 3979–3996.

[250] P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on deep learning architectures for time series forecasting," *International journal of neural systems*, vol. 31, no. 03, p. 2 130 001, 2021.

[251] R. Guthridge and V. Clinton-Lisell, "Evaluating the efficacy of virtual reality (vr) training devices for pilot training," *Journal of Aviation Technology and Engineering*, vol. 12, no. 2, p. 1, 2023.

[252] S. M. Asish, A. K. Kulshreshth, and C. W. Borst, "User identification utilizing minimal eye-gaze features in virtual reality applications," in *Virtual Worlds*, MDPI, vol. 1, 2022, pp. 42–61.

[253] Y. Yamato and S. Takahashi, "Gaze-based authentication method using graphical passwords featuring keypoints," in *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*, 2021, pp. 273–279.

[254] N. Sendhilnathan, A. S. Fernandes, M. J. Proulx, and T. R. Jonker, "Implicit gaze research for xr systems," *arXiv preprint arXiv:2405.13878*, 2024.

[255] I. Rigas, O. Komogortsev, and R. Shadmehr, "Biometric recognition via eye movements: Saccadic vigor and acceleration cues," *ACM Transactions on Applied Perception (TAP)*, vol. 13, no. 2, pp. 1–21, 2016.

[256] X. Wang and Y. Zhang, "Nod to auth: Fluent ar/vr authentication with user head-neck modeling," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–7.

[257] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead, "Unsure how to authenticate on your vr headset? come on, use your head!" In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, 2018, pp. 23–30.

[258] M. Sivasamy, V. Sastry, and N. Gopalan, "Vrcauth: Continuous authentication of users in virtual reality environment using head-movement," in *2020 5th international conference on communication and electronics systems (icces)*, IEEE, 2020, pp. 518–523.

[259] I. Olade, H.-N. Liang, C. Fleming, and C. Champion, "Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr)," in *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations*, 2020, pp. 45–52.

[260] A. Bhalla, I. Sluganovic, K. Krawiecka, and I. Martinovic, "Movear: Continuous biometric authentication for augmented reality headsets," in *Proceedings of the 7th acm on cyber-physical system security workshop*, 2021, pp. 41–52.

[261] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[262] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[263] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *2010 20th international conference on pattern recognition*, IEEE, 2010, pp. 3121–3124.

[264] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[265] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-sne effectively," *Distill*, vol. 1, no. 10, e2, 2016.

[266] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *Journal of neuroscience methods*, vol. 250, pp. 126–136, 2015.

[267] Y.-C. Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics & Epidemiology*, vol. 1, no. 1, pp. 161–187, 2017.

[268] C.-L. Deng, C.-Y. Tian, and S.-G. Kuai, "A combination of eye-gaze and head-gaze interactions improves efficiency and user experience in an object positioning task in virtual environments," *Applied Ergonomics*, vol. 103, p. 103 785, 2022.

[269] B. J. Hou, J. Newn, L. Sidenmark, A. Ahmad Khan, P. Bækgaard, and H. Gellersen, "Classifying head movements to separate head-gaze and head gestures as distinct modes of input," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–14.

[270] M. Garg, D. Ghosh, and P. M. Pradhan, "Multiscaled multi-head attention-based video transformer network for hand gesture recognition," *IEEE Signal Processing Letters*, vol. 30, pp. 80–84, 2023.

[271] Z. Hu, D. Haeufle, S. Schmitt, and A. Bulling, "Hoigaze: Gaze estimation during hand-object interactions in extended reality exploiting eye-hand-head coordination," in *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, 2025, pp. 1–10.

[272] Z. Wang, H. Wang, H. Yu, and F. Lu, "Interaction with gaze, gesture, and speech in a flexibly configurable augmented reality system," *IEEE transactions on human-machine systems*, vol. 51, no. 5, pp. 524–534, 2021.

[273] V. Nair *et al.*, "Unique identification of 50,000+ virtual reality users from head & hand motion data," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 895–910.

[274] I. Jarin, Y. Duan, R. Trimananda, H. Cui, S. Elmalaki, and A. Markopoulou, "Behavr: User identification based on vr sensor data," *arXiv preprint arXiv:2308.07304*, 2023.

[275] P. Kourtesis, E. Giatzoglou, P. Vorias, K. A. Gounari, E. Orfanidou, and C. Nega, "Examination of eye-tracking, head-gaze, and controller-based ray-casting in tmt-vr: Performance and usability across adulthood," *Multimodal Technologies and Interaction*, vol. 9, no. 8, p. 76, 2025.

[276] O. J. Nwobodo, G. S. Kuaban, K. Wereszczyński, and K. A. Cyran, "Enhancing learning in augmented reality (ar): A deep learning framework for predicting memory retention in ar environments," in *International Conference on Computational Science*, Springer, 2025, pp. 92–106.

[277] L. Zhang, T. Luczak, E. Smith, R. F. Burch, *et al.*, "Using microsoft hololens to improve memory recall in anatomy and physiology: A pilot study to examine the efficacy of using augmented reality in education," *Journal of Educational Technology Development and Exchange (JETDE)*, vol. 12, no. 1, p. 2, 2019.

[278] S. Gargrish, D. P. Kaur, A. Mantri, G. Singh, and B. Sharma, "Measuring effectiveness of augmented reality-based geometry learning assistant on memory retention abilities of the students in 3d geometry," *Computer Applications in Engineering Education*, vol. 29, no. 6, pp. 1811–1824, 2021.

[279] Z. Makhataeva, T. Akhmetov, and H. A. Varol, "Augmented-reality-based human memory enhancement using artificial intelligence," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 6, pp. 1048–1060, 2023.

[280] M. W. Eysenck and M. T. Keane, *Cognitive psychology: A student's handbook*. Psychology press, 2020.

[281] N. Unsworth and A. L. Miller, "An examination of individual differences in levels of processing," *Memory*, pp. 1–12, 2024.

[282] Y. F. Li, S. W. Lye, and Y. Rajamanickam, "Assessing attentive monitoring levels in dynamic environments through visual neuro-assisted approach," *Heliyon*, vol. 8, no. 3, 2022.

[283] D. Nolte *et al.*, "Combining eeg and eye-tracking in virtual reality: Obtaining fixation-onset event-related potentials and event-related spectral perturbations," *Attention, Perception, & Psychophysics*, pp. 1–21, 2024.

[284] C. K. Sahu, C. Young, and R. Rai, "Artificial intelligence (ai) in augmented reality (ar)-assisted manufacturing applications: A review," *International journal of production research*, vol. 59, no. 16, pp. 4903–4959, 2021.

[285] J. Kalafatovich, M. Lee, and S.-W. Lee, "Prediction of memory retrieval performance using ear-eeg signals," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 3363–3366.

[286] H. Gorin, J. Patel, Q. Qiu, A. Merians, S. Adamovich, and G. Fluet, "A review of the use of gaze and pupil metrics to assess mental workload in gamified and simulated sensorimotor tasks," *Sensors*, vol. 24, no. 6, p. 1759, 2024.

[287] N. Binetti, C. Harrison, A. Coutrot, A. Johnston, and I. Mareschal, "Pupil dilation as an index of preferred mutual gaze duration," *Royal Society Open Science*, vol. 3, no. 7, p. 160 086, 2016.

[288] M. Kolnes, A. Uusberg, and S. Nieuwenhuis, "Broadening of attention dilates the pupil," *Attention, Perception, & Psychophysics*, vol. 86, no. 1, pp. 146–158, 2024.

[289] B. Binias, D. Myszor, S. Binias, and K. A. Cyran, "Analysis of relation between brainwave activity and reaction time of short-haul pilots based on eeg data," *Sensors*, vol. 23, no. 14, p. 6470, 2023.

[290] B. Binias, D. Myszor, M. Niezabitowski, and K. A. Cyran, "Evaluation of alertness and mental fatigue among participants of simulated flight sessions," in *2016 17th International Carpathian Control Conference (ICCC)*, IEEE, 2016, pp. 76–81.

[291] C. Bybee, A. Belsten, and F. T. Sommer, "Cross-frequency coupling increases memory capacity in oscillatory neural networks," *arXiv preprint arXiv:2204.07163*, 2022.

[292] S. C. Wynn, C. D. Townsend, and E. Nyhus, "The role of theta and gamma oscillations in item memory, source memory, and memory confidence," *Psychophysiology*, e14602, 2024.

[293] Y. Suzuki, F. Wild, and E. Scanlon, "Measuring cognitive load in augmented reality with physiological methods: A systematic review," *Journal of Computer Assisted Learning*, vol. 40, no. 2, pp. 375–393, 2024.

[294] L.-M. Vortmann, L. Schwenke, and F. Putze, "Using brain activity patterns to differentiate real and virtual attended targets during augmented reality scenarios," *Information*, vol. 12, no. 6, p. 226, 2021.

[295] J. Shen, J. J. Dudley, and P. O. Kristensson, "Encode-store-retrieve: Augmenting human memory through language-encoded egocentric perception," in *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, 2024, pp. 923–931.

[296] S. Gargrish, A. Mantri, and D. P. Kaur, "Evaluation of memory retention among students using augmented reality based geometry learning assistant," *Education and Information Technologies*, vol. 27, no. 9, pp. 12 891–12 912, 2022.

[297] A. Mynick, A. Steel, A. Jayaraman, T. L. Botch, A. Burrows, and C. E. Robertson, "Memory-based predictions prime perceptual judgments across head turns in immersive, real-world scenes," *Current Biology*, vol. 35, no. 1, pp. 121–130, 2025.

[298] S. Maidenbaum, A. Patel, I. Garlin, and J. Jacobs, "Studying spatial memory in augmented and virtual reality," *bioRxiv*, p. 777 946, 2019.

[299] M. A. Shafique, A. Munir, and I. Latif, "Quantum computing: Circuits, algorithms, and applications," *IEEE Access*, 2024.

[300] Y. Y. Fein *et al.*, "Quantum superposition of molecules beyond 25 kda," *Nature Physics*, vol. 15, no. 12, pp. 1242–1245, 2019.

[301] M. Erhard, M. Krenn, and A. Zeilinger, "Advances in high-dimensional quantum entanglement," *Nature Reviews Physics*, vol. 2, no. 7, pp. 365–381, 2020.

[302] S. Gerlich *et al.*, "Quantum interference of large organic molecules," *Nature communications*, vol. 2, no. 1, p. 263, 2011.

[303] M. Nauenberg, "Max planck and the birth of the quantum hypothesis," *American Journal of Physics*, vol. 84, no. 9, pp. 709–720, 2016.

[304] J. S. Rigden, *Einstein 1905: the standard of greatness*. Harvard University Press, 2005.

[305] H. Kragh, *Niels Bohr and the quantum atom: The Bohr model of atomic structure 1913-1925*. OUP Oxford, 2012.

[306] A. Aspect and J. Villain, "The birth of wave mechanics (1923–1926)," *Comptes Rendus. Physique*, vol. 18, no. 9-10, pp. 583–585, 2017.

[307] D. J. Griffiths and D. F. Schroeter, *Introduction to quantum mechanics*. Cambridge university press, 2019.

[308] L. R. Jaroszewicz, K. A. Cyran, S. J. Klosowicz, and A. Mrozek, "Fiber optic sensor with a few-mode speckle pattern recognized by diffraction method," in *Interferometry'99: Techniques and Technologies*, SPIE, vol. 3744, 1999, pp. 386–393.

[309] K. A. Cyran, "Rough sets in feature extraction optimization of images obtained from intermodal interference in optical fiber," in *Interferometry'99: Techniques and Technologies*, SPIE, vol. 3744, 1999, pp. 241–252.

[310] Y. Zhao, H. Zhao, R.-q. Lv, and J. Zhao, "Review of optical fiber mach-zehnder interferometers with micro-cavity fabricated by femtosecond laser and sensing applications," *Optics and Lasers in Engineering*, vol. 117, pp. 7–20, 2019.

[311] A. Avdoshkin and F. K. Popov, "Extrinsic geometry of quantum states," *Physical Review B*, vol. 107, no. 24, p. 245 136, 2023.

[312] A. Romero-Osnaya, "Geometry of the quantum states of light in a mach-zehnder interferometer," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 698, 2016, p. 012 014.

[313] H. Verkade *et al.*, "Misconceptions as a trigger for enhancing student learning in higher education: A handbook for educators," 2017.

[314] W. Tarng and M.-C. Pei, "Application of virtual reality in learning quantum mechanics," *Applied Sciences*, vol. 13, no. 19, p. 10 618, 2023.

[315] W. Li, "Simulating quantum turing machine in augmented reality," in *Proceedings of the 2023 8th International Conference on Multimedia and Image Processing*, 2023, pp. 107–112.

[316] N. Suprapto, W. Nandyansah, and H. Mubarok, "An evaluation of the "picsar" research project: An augmented reality in physics learning," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 15, no. 10, pp. 113–125, 2020.

[317] P. Schlummer *et al.*, "Seeing the unseen—enhancing and evaluating undergraduate polarization experiments with interactive mixed-reality technology," *European Journal of Physics*, vol. 44, no. 6, p. 065 701, 2023.

[318] T. Weymuth and M. Reiher, "Immersive interactive quantum mechanics for teaching and learning chemistry," *arXiv preprint arXiv:2011.03256*, 2020.

[319] C. Maclean, A. Wolfe, S. Bhatti, A. Centino, and R. Ghannam, "Virtual and augmented reality as educational tools for modern quantum applications," in *2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, IEEE, 2022, pp. 1–4.

[320] R. L. Linn, *Measurement and assessment in teaching*. Pearson Education India, 2008.

Politechnika Śląska, Wydział Automatyki,
Elektroniki i Informatyki

Doctoral Thesis