1-1-2016

Towards System Agnostic Calibration of Optical See-Through Head-Mounted Displays for Augmented Reality

Kenneth R. Moser

Towards system agnostic calibration of optical see-through head-mounted displays for augmented reality

By

Kenneth R. Moser

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

August 2016

Towards system agnostic calibration of optical see-through

head-mounted displays for augmented reality

By

Kenneth R. Moser

Approved:

J. Edward Swan II
(Director of Dissertation)

Song Zhang
(Committee Member)

Derek T. Anderson
(Committee Member)

Magnus Axholt
(Committee Member)

T. J. Jankun-Kelly
(Graduate Coordinator)

Jason M. Keith
Dean
Bagley College of Engineering

Name: Kenneth R. Moser

Date of Degree: August 12, 2016

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: Dr. J. Edward Swan II

Title of Study: Towards system agnostic calibration of optical see-through head-mounted displays for augmented reality

Pages of Study: 142

Candidate for the Degree of Doctor of Philosophy

This dissertation examines the developments and progress of spatial calibration procedures for Optical See-Through (OST) Head-Mounted Display (HMD) devices for visual Augmented Reality (AR) applications. Rapid developments in commercial AR systems have created an explosion of OST device options for not only research and industrial purposes, but also the consumer market as well. This expansion in hardware availability is equally matched by a need for intuitive standardized calibration procedures that are not only easily completed by novice users, but which are also readily applicable across the largest range of hardware options. This demand for robust uniform calibration schemes is the driving motive behind the original contributions offered within this work.

A review of prior surveys and canonical description for AR and OST display developments is provided before narrowing the contextual scope to the research questions evolving within the calibration domain. Both established and state of the art calibration techniques

and their general implementations are explored, along with prior user study assessments and the prevailing evaluation metrics and practices employed within.

The original contributions begin with a user study evaluation comparing and contrasting the accuracy and precision of an established manual calibration method against a state of the art semi-automatic technique. This is the first formal evaluation of any non-manual approach and provides insight into the current usability limitations of present techniques and the complexities of next generation methods yet to be solved. The second study investigates the viability of a user-centric approach to OST HMD calibration through novel adaptation of manual calibration to consumer level hardware. Additional contributions describe the development of a complete demonstration application incorporating user-centric methods, a novel strategy for visualizing both calibration results and registration error from the user's perspective, as well as a robust intuitive presentation style for binocular manual calibration. The final study provides further investigation into the accuracy differences observed between user-centric and environment-centric methodologies.

The dissertation concludes with a summarization of the contribution outcomes and their impact on existing AR systems and research endeavors, as well as a short look ahead into future extensions and paths that continued calibration research should explore.

Key words: augmented reality, head mounted display, spatial calibration

DEDICATION

I would like to dedicate this work to my family, without whom none would have been possible. To my wife, Melissa, you are my greatest source of inspiration and motivation, and your patience and compromises during the final year of this work will never be forgotten. My parents, Ken and Marie, I owe the greatest thanks, for instilling within me the importance of education, learning, and seeking truth. Your struggles and sacrifices, especially during my childhood, mean the world to me. I am forever proud to be your son.

To God belongs all the glory for this work. He truly lead me all the way, at home and abroad. His providence provided me a wealth of funding and friendship. With Him, all things are possible.

ACKNOWLEDGEMENTS

I would like to express unmeasurable thanks to Dr. Christian Sandor, Dr. Hirokazu Kato, Dr. Goshiro Yamamoto, Dr. Takefumi Taketomi, and all of the students and staff at the Interactive Media Design Lab at the Nara Institute of Science and Technology. Without their warm welcome and support, this work would not have been possible. The memories of my adventures and time with you all in Japan will always be looked at fondly.

Deep appreciation is also extended to Dr. J. Edward Swan II, Dr. Stephen R. Ellis, Dr. Magnus Axholt, Dr. Mark Livingston, Sujan Anreddy, and the staff and colleagues of Computer Science at Mississippi State University. Your input, feedback, and encouragement during the early steps of my dissertation laid the foundation for this work and its continuum. It is my privilege to have continued Magnus' work and become a part of the growing community of researchers within Augmented and Virtual Reality technologies. A special thanks is also given to Dr. Yuta Itoh who has been a tremendous colleague and collaborator, with whom it has been an honor to work with.

Final acknowledgments are given to the various organizations that supported my doctoral work through fellowships, grants, and awards: The Naval Research Laboratory, The NASA Mississippi Space Grant Consortium, The National Science Foundation East Asia and Pacific Summer Institutes Fellowship, The Bagley Graduate Internship Program, The Japan Society for the Promotion of Science, The National Science Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

AR	*Augmented Reality*
AV	*Augmented Virtuality*
CG	*Computer Generated*
MR	*Mixed Reality*
RE	*Real Environment*
VC	*Virtuality Continuum*
VE	*Virtual Environment*
VR	*Virtual Reality*
DOF	*Degree(s) Of Freedom*
DRC	*Display Relative Calibration*
FOV	*Field of View*
GPU	*Graphics Processing Unit*
HMD	*Head-Mounted Display*
HUD	*Head-Up Display*
IMU	*Inertial Measurement Unit*
OST	*Optical See-Through*
RGB	*Red, Green, Blue*
VST	*Video See-Through*

NASA	*National Aeronautics and Space Administration*
SLAM	*Simultaneous Localization And Mapping*
PTAM	*Parallel Tracking And Mapping*
SPAAM	*Single Point Active Alignment Method*
INDICA	*Interaction Free Display Calibration*

CHAPTER 1

INTRODUCTION

Augmented Reality (AR) provides a powerful medium through which man is able to enhance and diminish his perception of the surrounding environment. The most common means for experiencing AR is through visual enhancements created by computer generated (CG) graphics overlayed onto the world. This virtual content may be crafted so as to appear to be floating in front of the user, which is useful for displaying menus, labels, and interface elements in a heads-up fashion, or augmentations may be designed so they seem to be a part of the world itself, locked into position relative to physical objects in the environment. The accessibility to AR content of either style has been largely supported by the production of low cost, compact, portable, personal computing devices equipped with an ever expanding array of sensors, cameras, and connectivity features. Head-worn displays, often referred to as head-mounted displays (HMDs), are particularly well suited for use with AR applications and offer inherent advantages over other device types.

Unlike hand-held systems, HMDs allow the user to maintain a constant hands-free view of AR content. The use of transparent, or Optical See-Through (OST), displays in particular offer a unique advantage over Video See-Through (VST) and Virtual Reality (VR) systems, by allowing a user to view both virtual AR imagery and the real world environ-

ment simultaneously from their own natural perspective. Unfortunately, current technology does not afford the ability to directly tap into the human visual system, precluding the applicability of well established computer vision based camera calibration methods, and necessitating the use of approximation techniques for estimating a rendering model to match the user's view through the device. As with all indirect approaches, OST calibration is prone to a variety of human and systemic error sources that may be mitigated, to an extent, but which inherently limit the efficacy of the end result. In addition, even though accuracy and precision are the highest priority, a delicate balance must also be maintained with the usability of a technique to enable access by non-expert and novice practitioners.

The intended goal of this dissertation is to provide an overview of the importance of accurate OST HMD calibration, the current obstacles and obstructions which limit the utility of available methods, and progress toward the development of standardized system agnostic principles and benchmarks to guide not only implementation but also evaluation practices for calibration techniques targeting next generation consumer devices. The most thorough compendium on the development of both OST display technologies and AR as a whole is provided within the doctoral thesis of Magnus Axholt [3]. This author highly encourages the reader interested in furthering their understanding of the underlying principles and components of any general AR system and the canonical progression of OST HMD technologies, to consider a thorough perusing of Magnus' work. It is not the explicit intent, nor the purpose, of this document to repeat the thoroughness of his compilation, but to update and build on the information therein.

The opening sections of this work provide a brief review and introduction to the defining characteristics of AR and HMD systems at large, before narrowing the scope of the content to the research areas and questions evolving within the domain of OST calibration. A concise exposition on the purpose, parameters, and methods for calibration of OST devices is provided, along with a brief review of prior user study assessments and the prevailing evaluation metrics and practices employed within. This review identifies the predominant trend of calibration results expected from environment-centric approaches and possible correlations and influences from human noise due to the necessitated manual alignments.

The second half of this document outlines the major contributions and additions that have been made to the general body of academic knowledge. Three major and three minor works are included in this exposition with emphasis placed on the motivations and goals to investigate the viability of user-centric manual calibration techniques for current and next generation OST HMD hardware, and the potential performance gains or decreases compared to environment-centric alternatives. Additionally, the expected benefits to future research endeavors is discussed for each. The concluding chapter summarizes the outcomes of the novel contributions and reiterates their impact on existing AR systems and research endeavors. A brief look ahead into future extensions and paths that continued calibration research should explore is offered in closing.

CHAPTER 2

AUGMENTING PERCEPTION

"Is all that we see or seem

But a dream within a dream?"

– Edgar Allan Poe, *A Dream Within a Dream, 1849*

I believe this excerpt, from one of Poe's final works, aptly depicts the peculiar and illusory nature of media within our present culture. The special effects and computer graphics industry, for example, have nearly perfected the art of visual manipulation, allowing for raw video footage of bizarrely dressed actors in front of green screens to be transformed into award winning cinematic experiences of super powered heroes traversing mysterious science fiction landscapes. We are also currently facing a renewed insurgence of so called *Virtual Reality* devices, which afford us an opportunity to delve into fantasy adventures of our own devising. These mediums are, of course, fashioned with the explicit intent of mentally removing us from our present reality. However, what if we were able to, somehow, merge the virtual world with our own? This concept of *Mixed Reality* may indeed one day lead us into a state of perception in which we are no longer able to distinguish between what is dream and what is genuine.

I feel that I would be somewhat remiss if I did not begin this dissertation by citing the now infamous statements from Ivan Sutherland regarding an *Ultimate Display*, "within which the computer can control the existence of matter" [138]. He continues describing the abilities of this display to make a chair "good enough to sit in" and a bullet "fatal". Based solely on these aspects, one might easily argue today that Sutherland's display may actually be more akin to a modern 3D printer, able to generate physical constructs from digital designs. It is relatively implicit, though, that Sutherland's intent is to illustrate a system that extends beyond simple object creation and actually depicts a mechanism for direct manipulation of the user's perception of reality. Even though our technology is still far behind the *Ultimate Display*, important strides are continually being made toward the creation of devices explicitly designed to arouse a variety of real sensations from virtual computer controlled stimuli.

2.1 The Virtuality Continuum

An almost self-evident contradiction arises from the term *Virtual Reality*, since we normally apply the notion of reality to objects, forces, and events within our corporeal existence. However, the general aim of VR is to induce those same physical and perceptual reactions one would normally encounter in reality through presentation of alternative, or synthetic, stimulations fully controlled by a computerized system. The experiences provided by pure VR, therefore, are not confined to the same immutable laws and limitations of our physical world, but allow for the delivery of truly novel sensations to the user. It may be the case though, and often is, that not every aspect of this alternative reality is able

to be controlled. Perhaps stimulation from both the real and virtual realities must combine, or mix, to produce the desired effect. Paul Milgram and Fumio Kishino actually provide a simplified classification scale, or *Virtuality Continuum*, denoting the common modalities through which real and virtual items may intermingle within a *Mixed Reality* system [91].

A simplified illustration of Milgram and Kishino's VC is provided in Figure 2.1. Purely virtual and purely real environments are naturally positioned at the extrema, with the hybrid MR environments placed along the interior portion. While classification of a MR system within this taxonomy may appear rather ambiguous, clear definitions do exist which direct categorization based on the modality of the environment and the augmenting or enhancement items.

Figure 2.1 The Virtuality Continuum as represented by Milgram and Kishino

2.1.1 Virtual Environment

A VE is conventionally defined as a fully immersive synthetic world [86]. The user's sensory experience is completely controlled by the system, with the quality of the simulation measured through the sense of physical presence within the VE. Shuemie et al. [124], as well as Bowman and McMahan [28], provide a more thorough exposition on this notion of *Presence* as it relates to VR. The construction of a convincing VE is non trivial, and it

is often the case that the implementation of certain features are governed by higher level design choices. The choice of display mechanism, for example, inherently narrows the feasible options for user locomotion within the VE. Those VR systems constructed around display walls [36, 106] or CAVE-like configurations [35] restrict the physical movements of the user to the bounds of the display, forcing the user to either walk in place [144] or use a treadmill or similar device [68, 101]. In contrast, the use of head worn displays [30], coupled with portable computing solutions, may allow full freedom of movement. Although tracking limitations often restrict usability to a finite volume, path planning and directed walking algorithms are able to simulate much larger spaces [152]. Burdea and Coiffet offer a more thorough survey of general VR technologies [29].

2.1.2 Augmented Virtuality

AV describes a particular class of VR systems in which features from the real world are purposefully and deliberately included in order to enhance, or augment, the context of the VE. The influence of the real world information, though, is still bounded by the rules and protocols particular to the VE. Simsarian and Akesson aptly illustrate this concept through their "Windows on the World" application, which implants video textures of real world objects into a VE [128]. A simpler example of an AV use case is the included visibility of a user's real hands within VR [49]. Even though the hands are visible, all gestures, motions, and actions are only as effective as the virtual experience itself permits them to be. AV is also beneficial for VEs that allow freedom of movement. Nahon et al. illustrate this utility in their VR setup by monitoring the real environment around the

user and revealing impediments before collision occurs [102]. Remote collaboration is also commonly facilitated through AV by abstracting the collaborative effort into a virtual space where each party and their contribution is visible to the other participants [20, 117, 118]. Largely though, AV systems are considered environment aware VR and therefore, to simplify future discussion, the VR denotation will also be considered to include AV instances as well.

2.1.3 Augmented Reality

The complement of AV, AR refers to the production of virtual information for the purpose of enhancing the user's perception of their real environment [14, 156]. Mackay describes AR as a unique interfacing paradigm between humans and computers [83], in which digital information is interwoven into the physical world to enrich the user's daily activities. The applications, domains, and benefits of AR are just as varied as that for computers themselves [16, 23, 151]. Doctors and medical professionals, for example, may utilize computer generated overlays to view ultrasound imagery directly on a patient's body [15]. Maintenance personnel can use world registered 3D models to guide repair and assembly tasks [44, 53, 125], and soldiers are able to create and share point of interest and situational awareness data with support teams across a battlefield [81, 162]. While these examples illustrate the use of digital information to add context and interest to the world, it can also be harnessed to hide or conceal features. This concept of *Diminished Reality* provides a powerful mechanism for removing undesirable or distracting components of an environment, which may include fiducial and computer vision markers or pieces of equipment and

furniture within a tracking space [34, 57, 90]. Whether for explication or camouflage, the accessibility of low cost conusmer devices, such as mobile phones, tablets, and wearble hardware, have made AR the most ubiquitous classification of MR [12, 56, 112, 155].

2.1.4 Real Environment

Simply for completeness, I will briefly discuss the general classification for an RE in regards to the VC. In the most general sense, an RE is the direct opposite of a VE. That is to say, an RE is fully non synthetic and composed entirely of the naturally occurring substances within our universe. This definition does not preclude computerized control of certain facilities within the environment, such as lighting, sound, or the movement of existing physical objects, such as by robots. However, items and energy within a RE are subject to all natural physical laws and the persistence of the environment is not dependent upon a user's presence or interaction. Stated more succinctly, the RE is the ground truth reference upon which all other virtual instances are measured and based.

Excluding the RE, all classifications across the VC share the common requirement for the presence of synthetic sensations, though albeit in varying amounts. A variety of actuators, chemicals, mechanisms, and novel hardware designs and approaches have been contrived to provide facilities for creating virtual stimulations intended to mimic each of a human's natural five senses. While some of these methods are more suitably applicable to certain system types, there is significant overlap in regards to the use of virtual stimulation for both VR and MR environments.

2.2 Sensory Augmentation

Creating truly immerssive and believable VEs or CG content for VR and AR relies heavily on the quality of the virtual sensations and stimulations used within the system. Naturally, as more senses are influenced by a particular application, the acceptability and trustworthiness of the synthetic items will also increase. Significant research efforts have been devoted to the development of stable, reliable, and robust sensory manipulation apparatus to address the ability to touch, taste, smell, hear, and see virtual objects, Figure 2.2.

(a) (b) (c) (d)

Figure 2.2 Augmented Reality modalities

(a) Multi-sensory system to augment taste perception [103]
(b) Self-contained olfactory system for outdoor localization using scents [159]
(c) Haptic device adapted for use in medical simulation [33]
(d) Use of CG for visualizing the perceived source of 3D sounds [38]

2.2.1 Haptics

Affording users the capacity to touch and feel virtual objects, as though they were physical, is a long standing goal for AR, MR, and VR development [70, 136], and is also the premiere requisite for Sutherland's *Ultimate Display*. Though size, mobility, accuracy, and calibration are limitations of current devices, frameworks to abstract and ease

integration have been proposed [40, 50], as well as compact glove designs [25] with less restriction on range of motion. Alternative, non mechanical, haptics is also possible by catering stimuli to other senses, particularly vision, to influence the perception of texture and plasticity [80]. Usable, cost effective, haptic systems have yet to be produced for the consumer market though, leaving the application domain largely restricted to the industrial sector. Figure 2.2 (c) shows a commercially available haptic device modified for use in medical simulation and training [33], and Srinivasan [133] provides a more exhaustive exposition on an array of haptic mechanisms.

2.2.2 Olfactory and Gustatory

Environment enhancement is also possible through the controlled delivery of odorants and tastants. Olfactory displays, as they are denoted in scholarly literature, incorporate pre-fabricated and loaded scents, usually in a liquid perfume-like form, with small directable air flow mechanisms [17, 159]. The ability to deliver the scent particles in varying quantities around the user creates the perception that the odors are emanating from virtual or real objects within the wold [161]. Figure 2.2 (b) shows an olfactory system designed for outdoor localization of scents. Multi-modal AR systems, such as that by Narumi et al. [103], extend the application of olfactory displays to gustatory responses by combining visual overlays onto food markers, with the intent that the combined visual and odorant augmentations will influence the user's perception of taste. A photograph of their system is shown in Figure 2.2 (a). The dependence on pre-fabricated scents greatly limits not only the num-

ber of odors able to be generated, but also the duration that the olfactory experiences can be maintained.

2.2.3 Audible

High definition audio is already common place in most entertainment venues, from movie cinemas to in-home surround sound for video gaming. Similarly, 3D sound generation allows for increased immersion and added realism of augmenting items [168]. Just like natural sounds from the world, these virtual sounds facilitate the localization of points of interest [131], further illustrated in 2.2 (d). Audible AR additionally provides a viable modality for deploying guidance systems intended for visually impaired or sightless users [73]. The advances in 3D sound generation for binaural headphone devices [38] shows further potential for application to current and future consumer level systems.

2.2.4 Vision

By far the most widely used, and well known, variety of synthetic stimulation is the use of CG graphics to add virtual content to the user's view of the environment. As previously noted, readily accessible mobile consumer devices provide low cost rendering solutions for VR and AR alike. Typically, virtual imagery is classified into one of two categories, statically registered *Head Up Display* or dynamically registered 3D content. HUDs provide an intuitive natural means for interaction, labeling, and general information retrieval [32, 54, 59, 129, 141]. The location of menus and other two dimensional interface elements remain fixed within the user's field of view, mimicking the layout style one might see on modern smartphones. 3D world registered content, however, renders CG graphics with the

intent to make the virtual items appear to be a part of the world [21, 22, 79]. This display style requires dynamic tracking and localization of the user's view to maintain the proper pose of augmented items as the user's gaze traverses across the world [119, 122, 165, 164].

As noted, haptic, olfactory, and gustatory technologies have yet to reach a viable level of performance, compactness, and reliability for use in consumer settings. Even though 3D audio advances are approaching mainstream, vision has reliably remained the most consistent medium for presentation of virtual content. The release of modern low cost consumer *Head-Mounted Display* devices have additionally begun replacing more traditional panel and projection based displays for the delivery of MR experiences.

2.3 Head-Mounted Displays and On-Going Research Interest

The growth in lightweight miniature display technology, heavily driven by the consumer mobile device market, has fueled an explosion in the availability of head-worn options. These devices bare only the slightest resemblance to Sutherland's early design [139], with many offering fully self-contained computing solutions, or at the least, an assortment of on-board sensors for measuring orientation, acceleration, and global positioning, as well as RGB and depth cameras for recording and identifying features within the environment. Ozan Cakmakci and Jannick Rolland offer a discussion of head-worn display types and trends [30], while Bernard Kress and Thad Starner offer a more focused exposition on HMDs specifically designed for the consumer market [78]. Despite the large variability in composition, feature sets, and styles however, HMD solutions can fundamentally be categorized, at a high level, into two groups: non-see-through and see-through [18].

Non-see-through displays, as the name implies, are completely opaque and fully, or at least partially, obstruct the wearer's view of their surroundings. This classification is synonymous with the standard stereotypes of huge bulky VR headsets portrayed in movies and television. Even though large industrial varieties are still in common use, slim low weight form factors are the current norm. Figure 2.3 shows several popular commercially available VR headsets. Although they are completely solid and opaque, it is possible to adapt these displays for use in AR applications by providing the wearer a view of the world through the video feed of cameras attached to the front of the device. Commonly referred to as *Video See-Through*, this approach affords a highly versatile and easily implemented mechanism for AR.

(a) (b) (c)

Figure 2.3 Popular commercial VR headsets

(a) Google Cardboard fitted to a standard Android based smart phone
(b) Samsung Gear VR headset by Oculus
(c) Oculus Rift Consumer Edition (v1)

Orlosky et al. aptly illustrate the potential of VST systems in their *modulAR* HMD configuration [109], which utilizes a varying assortment of mounted cameras to enhance the wearer's vision. For example, the feed from telephoto cameras provide a zoom onto

regions of interest, and fish-eye lenses enable a wider FOV than is possible with the naked eye alone. Aberrations and image distortion produced by the camera optics can, more often than not, be readily accommodated for and corrected through application of calibration methods, such as those developed by Tsai [147] and Zhang [166] for example. Calibration results also provide the viewing parameters enabling the graphics pipeline of VST AR systems to produce renderings of virtual objects that matches, almost perfectly, the perspective of the camera. Similarly, having direct access to the user's view, via the camera's video feed, natively allows the inclusion of on-line image processing and computer vision algorithms for position and orientation tracking of visible markers and natural features within the environment. The highly optimized performance of these tracking APIs, such as ARToolkit [72], is, without a doubt, the singular reason for the current prevalence and demand for AR applications on mobile and portable smart devices and hardware. There are, of course, a number of limitations and usability constraints inherent to camera based perspectives.

The most obvious deficiency in any VST system is the positional misalignment between the camera's image plane and the user's eye. Rigid camera fixations in binocular setups result in IPD mismatch, which influences the perception of distance due to improper stereoscopic depth cues [39, 150, 158]. Likewise, accommodation-convergence rivalry is unavoidable since the user's focal demand on the HMD screen remains fixed regardless of the eye's vergence angle. Although monocular and bi-ocular systems are able to circumvent these conditions through a single camera viewpoint, proprioceptive and vestibular discrepancies are an inescapable byproduct of VST in general. Hand-eye coordi-

nation tasks often require a great deal of kinesthetic training to adjust for the visual shift of the camera viewpoint [24, 134, 111]. Additionally, non-transparent displays, at large, are not well suited for situations with low fault tolerance, such as military combat situations, automobile or moving vehicle navigation, and delicate time sensitive medical procedures, where a device failure would make the wearer completely blind to their surroundings. In these scenarios, see-through displays offer the unique advantage of allowing the wearer to maintain a constant visual of the world regardless of the display state.

More commonly denoted as *Optical See-Through*, transparent display technology superimposes CG content directly onto the user's natural view of their environment. Optical combiners, such as prisms and partially silvered mirrors, coupled with compact lens arrays, for focusing and collimation, have been the most common approach for the design of OST devices, including the earliest models using CRT displays [139] to modern hardware releases using state of the art micro OLED screens. Figure 2.4 provides a closer view of a binocular OST HMD and its optical combiner. Alternative designs do exist though, which utilize high precision laser light to "paint" the CG imagery directly onto the user's eye. These *Retina Displays* not only reduce weight and compactness by removing the need for optical lens hardware, but are also able to provide correct accommodative cues by adjusting the focus of the laser as the image is drawn. As with their non-see-through counterparts, consumer models of both OST varieties are readily available on the market today, with next generation versions expected to be released in the near future. Of course, application development for OST HMDs is not without its own share of difficulties, especially in regard to calibration, which is far more user and system dependent compared to that for VST AR.

(a) (b)

Figure 2.4 OST HMD with partially silvered mirror combiner

(a) Epson Moverio BT-200 display with power and CPU unit
(b) **(Top)** Front view of the display lens and optical combiner
(b) **(Bottom)** Side view of the optical combiner within the display lens

Allowing users to continually view the world with their own eyes, and not through a VST system, means that the same calibration methods used to measure and match the rendering perspective of a scene camera are no longer able to be employed. Variations in head and facial structure between people, coupled with movement and shifting of the device during use, greatly diminishes the efficacy of a static view assumption, and results in the need for a per-user calibration methodology in order to maximize the accuracy and benefit of OST AR. Unfortunately, determining the location and view of the user's eye relative to the display screen is not a straight forward task, and little to no effort has been made by current device manufacturers to develop agreed upon standardized procedures applicable across hardware systems. Efforts from the research community have, nevertheless, given rise to a number of promising and viable calibration options, though thorough evaluation studies of the robustness and accuracy of these techniques have yet to be conducted. As to

date, a fundamental approach for applying ubiquitous system agnostic calibration has yet to be formally outlined.

CHAPTER 3

SPATIAL CALIBRATION OF OST HMDS

Calibration, in a general sense, refers to the process of measuring, modeling, or mapping relationships between two distinct quantites or sets. These sets may represent coordinate frames, colors, intensities, or perhaps even periods of time. The objective of OST HMD calibration is to compute the transformation of points from the 3 dimensional world space into the 2 dimensional pixel space of the display screen. This transformation is essentially encoded by the rasterization process of modern computer graphics pipelines and requires a description of the shapes and locations of the virtual objects to be rendered, usually stored in a vector graphics style format or vertex mesh, along with a mathematical model of the "camera" through which the virtual items are viewed. Since the camera in an OST device is actually the user's eye itself, a properly calibrated system will produce a rendered image that perfectly aligns with the user's view through the display screen

Consider the simplified illustration of the visual system formed by the eye and HMD optics in Figure 3.1 (a). The field of view of the user's gaze is driven by the relative position of the eye behind the display optics, which in turn determines the amount of visual angle over which virtual content is visible. Figure 3.1 (b) illustrates this system modeled as a pin-hole camera with an infinitely small aperture. This rendering volume, as employed in

most computer graphics libraries, produces a 2D perspective projection of objects within the frustum, with the field of view determined by the distance between the aperture and imaging plane. The goal of OST calibration is realized when the viewing frustums from 3.1 (a) and 3.1 (b) match, 3.1 (c). Of course, precisely modeling the user's viewing frustum is a highly complex problem with no direct mechanism for achieving an exact solution.

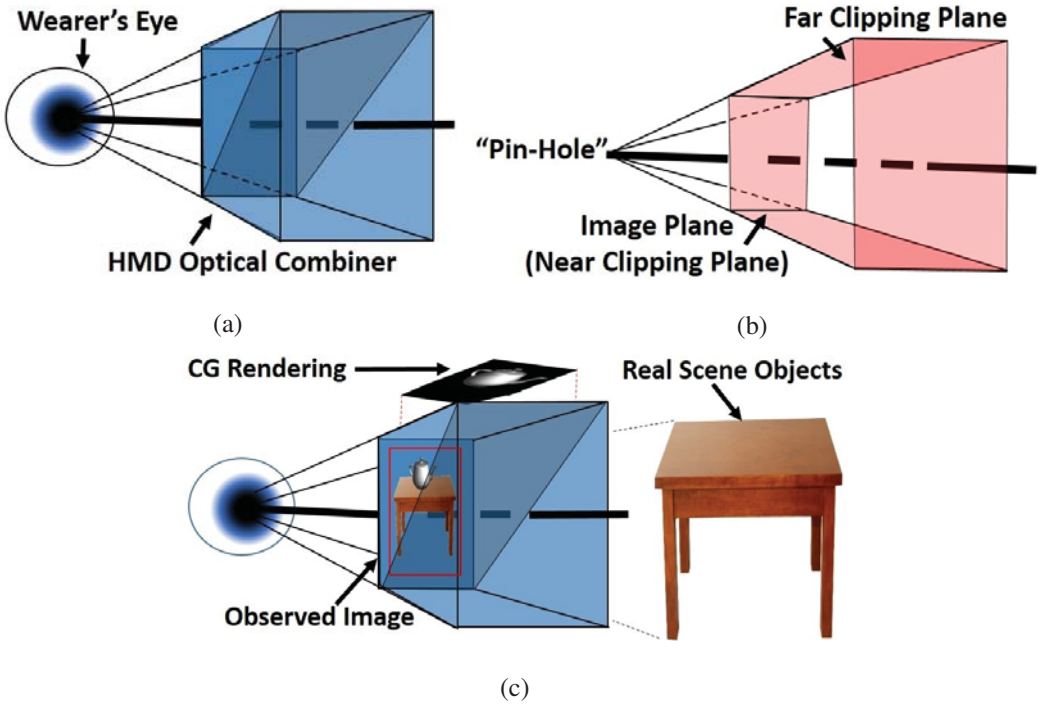


Figure 3.1 Illustrations of the real and virtual viewing frustums within OST HMD systems

(a) Depiction of the user's view through the optical combiner of an OST display
(b) Depiction of the virtual camera rendering volume in a standard graphics pipeline
(c) Illustration of a properly calibrated virtual camera

Even though it is not possible to see through the eyes of the user, there does exist a number of direct and two-phase calibration methodologies able to provide an approximate or estimated solution to the problem of calibration. However, there is little consensus

across these techniques on correct implementation strategies for reducing the effects of errors that arise from not only systemic sources but also the user as well.

3.1 Sources of Error

While calibration error can be described conceptually as a mismatch between the viewing frusta of the eye and virtual camera, this error manifests itself literally as a horizontal or vertical shift, rotation, or scaling offset of the CG geometry in the 2D image shown on the display. These visual errors, commonly referred to as registration errors, can be likened to the manifesting symptoms of a much larger calibration ailment. They are simply the result of the problem, and as such, attempting to correct these screen errors directly would be to ignore the underlying causes themselves. While it is not possible to directly map an error source to a particular type of registration error, since an infinite combination of differing error sources may create identical visual errors, a system designer knowledgeable in error contributors will be more readily equipped to identify potential problems and minimize or ameliorate error sources during the design stage.

Richard Holloway assembles the causality of registration error [60] into four basic categories: acquisition, tracking, display, and viewing errors. Viewing error in this case refers to an incorrect assumption of the user's eye point, which is addressed within the discussion of display error. It is worth noting that the contribution from each of these sources is not equivalent in every AR system and is highly dependent on the structure, interdependence, and rigor with which an application is designed.

3.1.1 Modeling

Acquisition, also referred to as modeling, error refers to the lack of resolution in measurements of the environment used to build internal geometry references. For example, situational awareness AR applications may wish to include wire-frame outlines of buildings and windows to identify points of interest for the wearer [81]. The positional fidelity of the wire-frame overlay is determined by how accurately the available computer model matches the actual dimensions of the building. Perhaps in this example, the precision of the computer model would still be visibly acceptable with a minimum resolution of 1 to 50 centimeters. Would this same dimensional limit continue to suffice in a surgical AR system designed to overlay a wire-frame onto major arteries for the surgeon to avoid? I believe most rational persons would agree that a precision of millimeters or even less would be required in this instance. This notion of an acceptable accuracy threshold is an important consideration for application designers especially when resources allocated to creating an environment model are limited.

The increasing availability of depth cameras, sometimes referred to as IR time of flight sensors, especially at the consumer level, has greatly diminished the burden on developers requiring a model of the user's environment prior to run time. The Microsoft Kinect and Asus Xtion sensors, for example, are capable of scanning, creating a mesh, and color mapping their surroundings at run time [31, 55, 105]. This technology will purportedly be an integral feature of next generation OST HMDs, including the Microsoft Hololens and Meta 2 devices, and will soon be available on mobile platforms as well. Current systems not equipped with depth cameras, however, may still be able to perform real-time envi-

ronment modeling using standard RGB cameras. *Simultaneous Localization And Mapping* procedures, originally intended for computer vision based robot guidance systems, utilize successive camera images to estimate the relative distances between feature points within the environment [37, 92, 146]. Tracking the movement of known feature points across an image series provides a measure of parallax through which 3D relationships can be extrapolated. Variations of the SLAM methodology, including PTAM [74], LSD-SLAM [42], and ORB-SLAM [100], include assumptions about the scale of the environment, distance ranges within the image space, or expand feature sampling across stereo-camera pairs. GPU accelerated algorithms are also extending the capabilities of these algorithms for use on mobile smart devices [75].

The accuracy thresholds on these camera based modeling algorithms is extremely hardware dependent, and the only means for determining the error resolution of a particular implementation is to have a ground truth model for comparison, which would, of course, be contradictory to the purpose of the algorithm itself. Nonetheless, the benefits of mapping the world at run time allows AR application developers to maintain an agnostic approach with regard to the user's environment, and once the static layout of the surroundings is known, it is straightforward to detect which portion is in view.

3.1.2 Tracking

Tracking, also denoted as localization, refers to the determination of an object's 6 DOF pose within a particular coordinate frame. An exhaustive explication of general tracking types is beyond the scope of this work, though referral to Ronald Azuma [11], Eric

Foxlin [45], and Bostanci et al. [27] will provide a survey of the basic requirements and most widely used modalities for AR tracking within indoor and outdoor environments. In general, any tracking mechanism can be classified at a higher level as being either *outside-in* or *inside-out*. These labels more laconically denote whether the sensors are fixed and the objects are in motion (outside-in), or if the objects are fixed and the sensors move (inside-out). The IR optical tracker shown in Figure 3.2 (a) provides outside-in tracking, since the camera sensors are rigidly fixed in the environment and provide pose estimations for the movable targets within the tracking volume. Figure 3.2 (b) illustrates an inside-out tracking structure where the graphical targets are secured in place and the camera sensor moves around the environment. Even though it is possible to measure both the position and orientation of an object using either arrangement, it is not uncommon for readings from multiple sensor types to be combined and aggregated for enhanced robustness and resilience to errors within a single tracking source.

(a) (b)

Figure 3.2 Example tracking systems

(a) Illustration of an outside-in tracking system using four Optitrack IR cameras to measure the location of objects within the visual volume
(b) Depiction of a hybrid inside-out tracking system with the pose of a user determined by environmentally located fiducial markers tracked via an on-board camera [113]

A *Sensor Fusion* [163, 123, 58, 76] approach is a best practice especially when integrated IMU hardware is available within the system. A computer vision based primary tracker, such as one of the SLAM approaches noted in section 3.1.1, will experience complete failure when a moving obstruction, such as a user's hand, passes in front of the camera. Readings from an accelerator and gyroscope, in this instance though, would allow the program to predict the motion of the device until a visual marker is once again in view. Sensor fusion systems, as with single source tracking types, are still only capable of providing pose estimates over discrete time sequences with a finite resolution. Improper synchronization of tracking data with rendering frames results in a visible lag or latency of the virtual objects' position as the user's view moves about the world. The effects of this temporal error naturally effect usability as any task requiring precise interaction will be bounded by the update rate [69, 41, 89, 1]. Likewise, simulator sickness [135] may be induced from miscorrelation between visual, vestibular, and motor stimuli.

Error in positional accuracy and precision is also exhibited by every tracking system and is a by product of either measurement resolution, range limitations, data noise, interference, or any combination there of. As discussed in section 3.1.1, the resolution of a tracking system may preclude its applicability to certain domains and situations. The range over which tracking data maintains reliability may be subject to system specific factors, such as visual obstructions, and will often degrade as measurements proceed to the boundary of the tracking volume. Noise and interference are distinguished by the stability and predictability of errors, with noise more precisely describing a continuous jitter within a predictable range and interference denoting a corruption in data due to an external in-

fluence. Teather et al. [143] expound in more depth on the effects of latency and jitter on virtual objects.

3.1.3 Display

The ultimate quality of AR registration can only be determined once virtual content is rendered onto the HMD. Ideally, all visible registration errors would be limited to the effects of modeling and tracking as discussed prior. Realistically though, distortion, warping, and shifts perceived in augmentations is partially a result of refraction due to aberrations and defects in the optical components of the display. This includes the optical combiner, lenses, and perhaps even the imaging element itself. Fortunately, optics is an extensive and well studied domain with a plethora of available strategies, mechanisms, methods, and procedures for addressing display issues.

Identically to camera calibration [147, 166, 126, 157], the distortion from OST lenses can be modeled through tangential and radial components [137] as well as through non-parametric regression methods [116, 51], and procedures have been proposed for application to HUD systems [154]. These standard correction schemes are able to provide a reasonable correction for most systems, especially for those with very minimal distortion throughout or concentrated to the extents of the FOV. Improvements based on camera calibration, though, are actually only able to provide distortion correction for a single viewpoint through the optics. Since the refraction pattern will not remain constant as the user's eye moves relative to the screen, a different approach is required for optimal correction. Itoh et al. [66] model the collective distortion a user experiences as a 4D light field map-

ping. This enables a per user correction, provided that the location of the user's eye can be determined with an appropriate amount of accuracy.

Distortion and optical aberrations are not the only modality of registration error generated by display hardware though. As described earlier, the user's view is fundamentally presumed to mimic that of a pin-hole camera system, where the imaging plane is perfectly perpendicular to the viewing direction and the frustum is symmetric. Given the wide array of HMD hardware and imaging mechanisms, many of which assume a static IPD across users, these assumptions are most often not satisfied, and the user's view through the display produces an off-axis or asymmetric viewing frustum [167, 121]. Viewing aberations may likewise result from an angled or canted imaging plane. In this scenario, the cant may be converging or diverging with regards to the intersection of the left and right eye views (for stereographic systems), leading to incorrect vergence angles of users' eyes producing erroneous perception of depth of virtual objects [39, 132].

Finally, display latency, similar to tracker latency, will also produce temporal registration error. Display latency refers to the time required for the final rendered image to appear on the screen and is a factor of the imaging system and data channel. Nearly all current OST HMD hardware sets utilize a wired connection for delivery of the video signal, though the current trend is leading production for complete computing solutions with on-board rendering capabilities. In either case, noticeable latency in the display image is unavoidable, simply because the user's view of the world is updated continuously and non-uniformly, while that of the display screen is discrete and instantaneous. Therefore, movement of the user's head and eyes between frame updates will naturally cause misalignment

between the current frame and the visible world. Post rendering image warping [85, 130] is gaining popularity as a mitigation strategy for simulating continuous imagery, by warping the current frame as a function of the viewpoint direction, position, and orientation of the user during each rendering cycle. This method requires additional computing resources, though hardware based implementations are already in use on several commercial head worn devices, including the Oculus Rift and HTC Vive.

3.2 Calibration Methods for OST HMDs

As discussed in the opening portion of this chapter, the goal of OST HMD calibration is to correctly model the user's view through the display, by matching the viewing frustum of the rendering engine to that of the eye. The resulting image projection is based on the pin-hole camera model described by an infinitely small aperture, through which incoming light rays pass, and an imaging plane intersected by the rays onto which the image is formed. In the physical world, light passing through the aperture may originate from an infinite distance. This assumption is an impossibility for computer graphics pipelines however, which are limited by memory and computational precision, allowing rendering engines to only model a discrete volume of space. The extent of visible objects in rendering space is represented by the addition of clipping planes, though only the "far" plane is a requirement. The final image, therefore, is effectively produced by a coordinate transformation of the virtual objects from the 3D rendering volume to the 2D image space.

3.2.1 What is the Projection Matrix?

Coordinate transformations, in computer graphics, are expressed algebraically through matrix operations. The rendering perspective projection operation is no different, and the 3D world to 2D screen transformation is encoded as a 3×4 matrix, or 4×4 when converted to use homogeneous device coordinates. Equation (3.1) provides this camera projection using the notation from Tuceryan and Navab [148]. All calibration methods, therefore, must be able to produce this projection matrix by either solving for all of the matrix components at once, or by systematically determining the parameters in stages.

$$T_{camera} = \begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \\ \begin{bmatrix} a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \end{bmatrix} \tag{3.1}$$

3.2.1.1 Intrinsic Display Components

The T_{camera} projection transformation describes not only the intrinsic camera perspective but also the extrinsic location of the camera in the relative coordinate frame. Equation (3.2) provides the relationship between T_{camera} and its intrinsic T_{proj} and extrinsic T_{pose} components.

$$T_{camera} = T_{proj} * T_{pose} \tag{3.2}$$

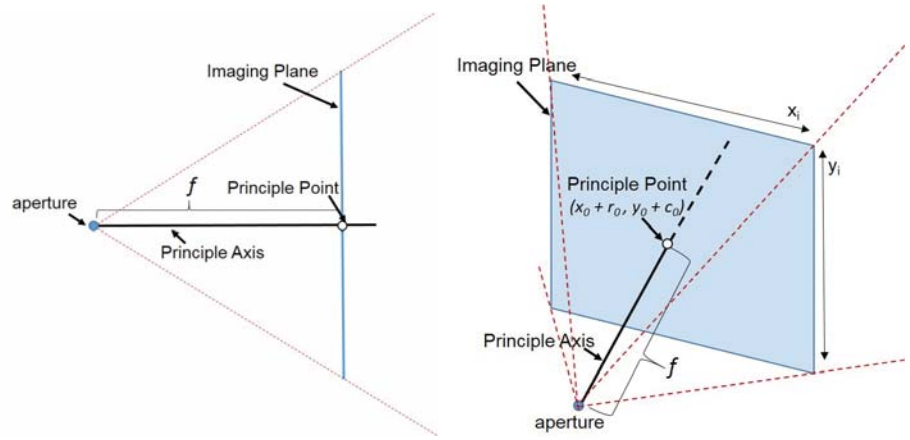
The intrinsic component matrix, as the labeling denotes, defines the projection transformation from 3D to 2D coordinate spaces. The elements of this matrix describe the properties of the pin-hole camera and its derivation is well described in a plethora of aca-

demic texts and research publications [52, 43, 63, 82, 148, 153]. Readers desiring to gain a complete and thorough understanding of the physical and mathematical principles behind projection, transformation, or computer graphics in general are encouraged to read the cited publications, but for clarity sake, a brief review of the parameters of the projection matrix will follow.

$$T_{proj} = \begin{bmatrix} \begin{bmatrix} f_u & \tau & r_0 & 0 \\ 0 & f_v & c_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{bmatrix} \tag{3.3}$$

The parameters of T_{proj} from Equation (3.3) are derived directly from the pin-hole camera model from Figure 3.1 (b). A simplified 2D side and 3D rear view of the pin-hole virtual camera is shown in Figure 3.3 (a) and (b). The focal distance, f denotes the distance between the imaging plane and the aperture of the camera. In the ideal pin-hole camera model the f_u and f_v components from Equation (3.3) are identical, the pixels of the image are perfectly square. While, the rendering engine in computer graphics is an ideal pin-hole camera, in physical implementations these values may be unequal as a result of distortion, imperfections on the imaging plane, non-uniform image scale, etc., in which case an alternative model using a single focal length value and the image aspect ratio may be more appropriate [115]. The "principle axis" lies perpendicular to the imaging plane and extends to the aperture. The intersection of the principle axis and the imaging plane occurs at the "principle point". Ideally, the principle point would occur at the origin of the image coordinate system. However, when this is not the case, the parameters r_0 and c_0

represent the offset from the origin. The remaining value τ is not shown in Figure 3.3. τ represents a skew factor when the axes of the image plane are not orthogonal, which would produce an image plane resembling a parallelogram instead of a rectangle or square.



(a) (b)

Figure 3.3 Illustration of pin-hole camera projection

(a) Side-view of pin-hole camera projection system
(b) 3D view of pin-hole camera projection system

3.2.1.2 Extrinsic User Components

When the camera is located at, and is orthogonal to, the origin of the 3D coordinate space, then the transformation of objects into the camera frame of reference is implicit. However, should the camera move to another viewing location in the world, as is often the case, then an extrinsic transformation is required to transform the coordinates of the objects in the world into the camera frame. This transform is the T_{pose} component of Equation (3.2), whose matrix form is provided in Equation (3.4)

$$T_{proj} = \begin{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \end{bmatrix} \tag{3.4}$$

The $r_{11} - r_{33}$ components describe the rotation of the camera with respect to the world coordinate axes and the $t_1 - t_3$ components denote the translational offset from the origin along the X, Y, Z cardinal directions. This transformation, with respect to OST HMD calibration, represents the transformation of the eye relative to the tracked coordinate frame of the HMD. As discussed in section 3.1.3, algorithms from the computer vision domain are able to derive the intrinsic values for a given eye point. Unfortunately, the location of the user's optical center, or alternatively the nodal point, is not easily determined at run-time. Nonetheless, given the extrinsic and intrinsic parameters, calculation of the 12 values in the final camera projection matrix T_{camera} in Equation (3.1) is through simple matrix multiplication.

3.2.2 Manual Approaches

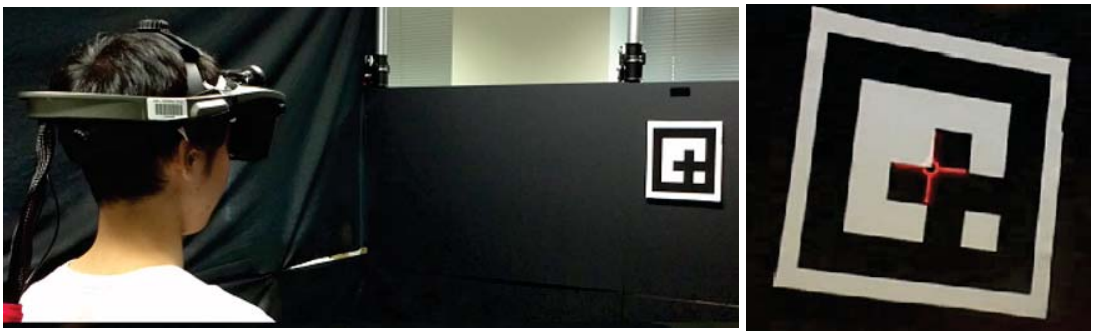
Since it is not possible to access the user's view through the display, OST HMD calibration must use approximation methods to estimate the parameters of the projection matrix. Even though these methods can not see through the user's eyes, it is still possible to obtain usable measurements based on feedback from the user about what they are able to observe. Initial calibration modalities, for example, adapted computer vision camera calibration mechanisms, which utilize pixel to world correspondences for determining the

viewing parameters. Instead of obtaining all correspondences at once, as would be possible in an image captured from a camera, these bore-sighting strategies instead record each correspondence in a sequence by having users manually adjust the location of on-screen reticles to align with a number of specific target points, of known locations, in the environment [13, 32, 77]. This schema forces a number of requirements, including placement of the HMD such that the user's view is perpendicular to the display screen and that the user is able to reliably align the on-screen indicator with a high level of precision. In order to satisfy these conditions, the user's head must be rigidly secured, preventing movements which may shift the display screen or disrupt the alignment process. Inhibition of user movement makes this methodology not only uncomfortable and tedious, but also impractical for use outside of a laboratory setting. Successive adaptations though have enabled the relaxation of the fixation constraint by affording a compromise with the other requirements as well.

3.2.2.1 Single Point Active Alignment Method

Mihran Tuceryan and Nassir Navab published a description for a revised manual calibration method at the 2000 ISAR symposium. Their procedure removed the necessity for head fixation allowing users to perform screen to world correspondence alignments with full freedom of motion [148]. The denotation *Single Point Active Alignment Method* (SPAAM) succinctly describes the process during which a user actively aligns a sequence of on-screen points to a single target location in the world. Unlike bore-sighting, SPAAM allows the same world target point to be reused for all correspondence pairs. This is pos-

sible due to the relaxed mobility constraints which now allow the user to freely move and rotate their head to make the alignments. Figure 3.4 illustrates a user's actions during a normal SPAAM calibration. A more thorough explanation of the mathematical solution behind SPAAM, which directly solves for the final 3×4 projection matrix using the 2D–3D correspondence pairs captured by the user, is provided in the original work [148].



(a) (b)

Figure 3.4 View of user performing a SPAAM calibration procedure

(a) Users move their body and head to align on-screen indicators with tracked world points
(b) View through the HMD of a screen–world alignment

Subsequent developments by Yakup Genc et al. [6] extend the basic SPAAM implementation to binocular OST HMDs, allowing for the simultaneous calibration of both the left and right eye views together. Similar to the monocular base case, Stereo SPAAM requires alignment between world and screen points with the latter exploiting stereoscopic depth cues afforded by binocular displays to create 3D virtual points, perceived at depth, in contrast to standard 2D reticles. Coupling the calibration of both eyes naturally decreases the completion time burden and explicitly provides cues necessary to ensure non-planar alignment recordings. Arthur Tang, Ji Zhou, and Charles Owen [142] discuss the benefits

of depth varied correspondence pairs for SPAAM calibration. Further efforts to reduce the user's workload, but maintain a user-centric approach, employ hand held tracking markers for alignment based calibration [72, 108]. Moving the marker around the field of view, mimicking the distance variation of Depth SPAAM, yields viable accuracy results and reduces user movement since the entire procedure may be conducted while seated. Even though this method offers a usable environment agnostic procedure, like other manual calibrations it too ultimately requires the possession of external alignment targets to proceed.

In addition to usability, strategies to decouple the determination of the intrinsic and extrinsic properties, and thus improve the robustness to error influences, have also generated a number of calibration variants. An example of such a strategy is the "Easy SPAAM" approach [47, 104] intended to optimize recalibration by adjusting an existing projection matrix with an updated location of the user's eye. This new position may be obtained in a variety of ways, though triangulation through the familiar user driven 2D–3D alignment procedure is often the most applicable. This recycled results approach, while still bounded by the accuracy of the existing calibration data, significantly reduces the number of screen to world correspondences required to recalibrate an HMD and also isolates the impact of further alignment error to the new extrinsic, eye location, measures. Alternative strategies require a two step process to completely isolate intrinsic and extrinsic errors.

3.2.2.2 Display Relative Calibration

Charles Owen, Ji Zhou, Arthur Tang, and Fan Xiao [110] outline a procedural methodology for measuring the display dependent properties of the viewing matrix completely

independently from the extrinsic user specific components. The first phase of their *Display Relative Calibration* (DRC) is conducted off-line, and leverages existing camera and projector based calibration procedures [160, 64] utilizing image processing and computer vision techniques for measuring not only the focal length of the HMD screen, but also the pixel scaling factor and apparent screen depth. Theoretically, these parameters will remain constant across production releases for each HMD model, allowing for the manufacturers themselves or astute research groups to measure and publish display data for direct use by system designers and developers. The remaining extrinsic values, which describe the placement of the user's eye within the head-mount, must then be determined at run-time.

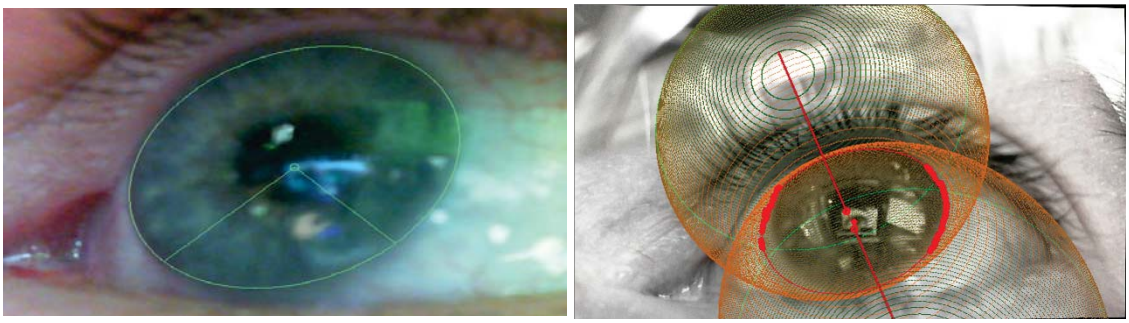
On-line triangulation methods, such as that discussed for Easy SPAAM, are of course viable options for obtaining the eye position data needed to complete the calibration. The calibration time for the user, employing the DRC approach, would therefore never appear any longer than the recalibration stage of the recycled SPAAM methodology. Unfortunately, reliance on user alignments for extrinsic values ultimately means that the intrinsic properties reflect a far greater robustness to measurement error. Ideally, complimentary techniques for measuring eye location will be able to provide a consistent and measurable accuracy tolerance, as well as incorporate procedures that not only reduce the initial calibration time requirements at system start-up, but that are also able to update extrinsic measures throughout the run-time cycle.

3.2.3 Semi-Automatic Approaches

The utility of user alignment is naturally necessitated by the limited information attainable by current generation HMD hardware. Inclusion of additional sensor devices, however, broadens the availability of relevant data for use in calibration. Of particular note is the growing accessibility of miniature high-definition cameras, which is slowly driving the development and release of low-cost eye-tracking systems [71]. The potential uses for eye-tracking within HMD hardware is varied and often focuses on hands free interaction and selection of virtual content [19]. However, Yuta Itoh and Gudrun Klinker have applied the notion of eye imaging to enhance the on-line phase of the DRC methodology.

Interaction Free Display Calibration (INDICA), as described in the original work [64], combines developments in eye recognition and positioning from the computer vision domain with low cost imaging solutions for HMD devices to fully automate the measurement of extrinsic parameters. At its premiere, the first INDICA ready system leveraged the iris detection algorithm described by Lech Swirski [140] to identify a user's eye within a captured RGB image. 3D localization of the eye center is then performed based on a general physical eye model describing the standard iris diameter and expected eye radius and corresponding center position, according to the process outlined by Christian Nitschke et al. [107]. Annotations reflecting the iris detection and 3D localization processes within example eye images are provided in Figure 3.5. While inherently providing extrinsic values for a DRC approach, applicability of the INDICA methodology also naturally carries over to Easy SPAAM and similar data recycling procedures as well. Subsequent improvements to localization accuracy through eye imaging has been proposed by Alexander Plopski et

al. [114], by exchanging iris detection with recognition of known patterns visibly reflected on the user's cornea.



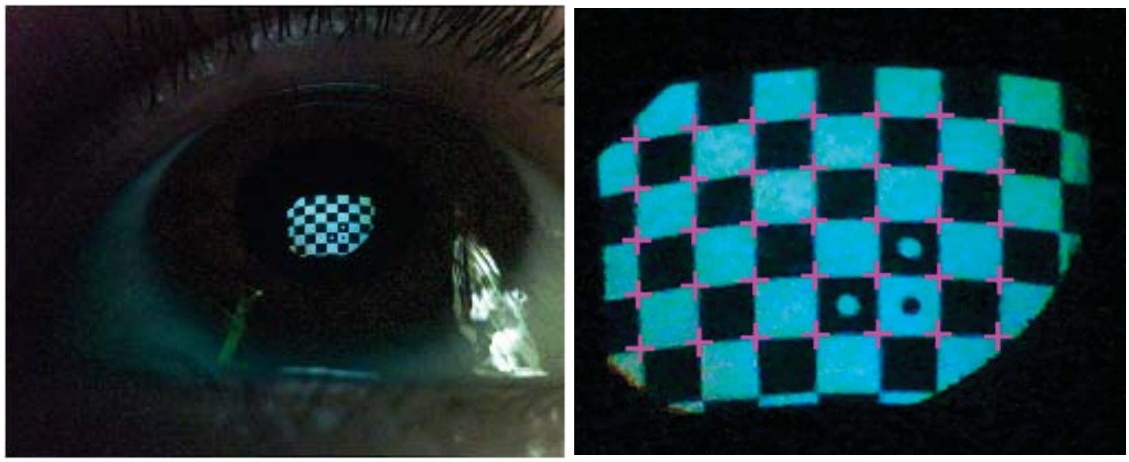
(a) (b)

Figure 3.5 3D eye location through corneal tracking

(a) Processed eye image showing the fitted elipse to the detected iris
(b) The fitted elipse from (a) is projected into a pair of symmetric spheroids. The 3D location of the eye, relative to the camera, is taken as the center point of the forward facing spheroid

Their *Corneal Imaging Calibration* (CIC) compares the resulting distortion of a detected pattern reflected on the eye's surface, Figure 3.6, against the undistorted ground truth image shown on the HMD. Relying on a uniform eye surface model, the observed warping of the reflected image is used to predict not only the position of the eye relative to the screen, but also the gaze orientation as well. Preliminary metrics presented in the original work show higher accuracy and precision estimates over the iris detection scheme. However, the current computational complexity, and reliance on random sampling and consensus (RANSAC) strategy for model fitting, has yet to realize the solution for use at interactive on-line rates. Likewise, both the INDICA and CIC processes require access to captured images of the user's eye at run-time. This requisite, of course, imposes an additional hardware constraint, that of rigidly mounted eye-tracking cameras, which must

be accommodated by either the HMD manufacturer directly or the system designer during development. As suitable imaging solutions and related components are not yet standard in current consumer or professional HMD products, and third-party options mostly attainable only at extreme cost, the implementation burden for both hardware consignment and software integration falls to the researcher post purchase.



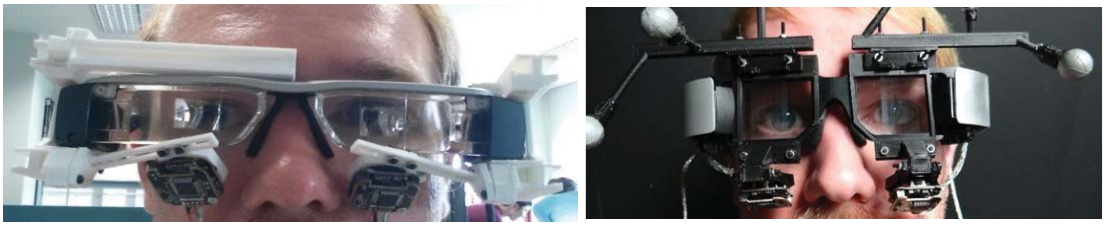
(a) (b)

Figure 3.6 Reflected pattern on the eye's surface

(a) Large view of the wearer's eye with corneal reflection
(b) Closer view of the corneal reflection.

Even though access to reasonably cost effective 3D printing and camera technologies may allow enterprising investigators to overcome the barriers of hardware integration, development of imaging attachments must be addressed for each possible HMD model available. Figure 3.7 provides photos of custom camera mountings designed and created by the author for two commercially available OST HMD systems. This obligation to craft tertiary camera systems, is an impractical consideration for wide-spread application of these automatic approaches to current hardware options. Additionally, the performance of

both INIDCA and CIC has yet to be formally evaluated in comparison to standard manual calibration approaches on identical hardware systems. Thorough objective and subjective assessment is essential to not only verify correctness in an active setting but also further quantify the utility and ease of use for each approach with regard to both developers and novice practitioners alike.



(a) (b)

Figure 3.7 Custom eye camera mountings

(a) 3D printed mountings for Epson Moverio BT 200
(b) Lumus DK-32 fitted with custom camera mounts

3.3 Evaluating OST HMD Calibration

Since the fundamental goal of OST HMD calibration is to correctly model the user's viewing perspective, the ultimate measure of calibration quality naturally derives from the degree to which registration error is minimized. Unfortunately, as discussed in section 3.1, visible registration error is purely the symptomatic result of more complex underlying errors from tracking, modeling, and display inaccuracies. Therefore, evaluating calibration based purely on the apparent registration quality of a single or diminutive set of viewpoints will not provide any substantial guarantee that the observed level of accuracy will be maintained across the entire spectrum of possible viewpoints accessible to the user.

Consequently, alternative less direct metrics for wholistically appraising calibration must be used and acquired through measurable objective quantification and subjective qualitative responses.

3.3.1 Objective Metrics

The benefit of any quantitative measure is separation from subjective bias and personal notions of quality and scale between individual users. Ideal objective measures will facilitate comparison across differing design implementations by providing a quantity with equivalent meaning and measurable accuracy irregardless of physical setup. The extrinsic component of the calibration result inherently provides quantifiable information with consistent implications.

One comparatively uniform measure is the positioning of the user's eye within the head-mounted device. Since the tracked coordinate frame of the HMD is established prior to any calibration procedure, a general region for the expected locations of the user's eyes within the display device can also be established prior to use. Any deviations between the expected localities and the modeled values ascertained from calibration results will therefore expose explicit deficiencies in the extrinsic component of the projection matrix. Secondary metrics are also attainable using the eye position measures by considering the left and right eye locations mutually. These binocular disparity values describe the relative differences in eye locations along the three major axis of the head.

Inter pupillary distance (IPD) describes the horizontal separation between eye centers, Figure 3.8 (a), and similar to the general extrinsic transformation, can be compared against

a known ground-truth value measured for each user prior to calibration. Quantifying any IPD mismatch within the system provides a reference for predicting possible misperception of depth in virtual content [39] during use. Ground-truth measures for the lateral, Figure 3.8 (b), and vertical, Figure 3.8 (c), eye separations are not as straightforward. However, comparable measures across systems is possible as long as consistent and reasonable presumptions of symmetry are maintained. While independent estimation of the extrinsic parameters is possible through two step DRC techniques, such as INDICA and CIC, manual calibration methods, including SPAAM, couple the estimation of both the intrinsic and extrinsic properties together. This concurrent estimation makes isolation of any extrinsic specific error an impossibility due to innate systemic influences from the intrinsic parameters of the display. Therefore, SPAAM-like calibrations often also consider reprojection error as an additional quality metric for objective evaluation.

(a) (b) (c)

Figure 3.8 Illustrations for binocular disparity metrics along the three cardinal directions

(a) Horizontal (left–right) disparity, IPD
(b) Lateral (forward–back) disparity
(c) Vertical (up–down) disparity

As described in section 3.2.2.1, the SPAAM calibration procedure utilizes a series of user driven screen to world correspondence pairs to solve for the 12 values of the user

projection matrix directly. Reprojection refers to the process of applying the projection matrix result to transform the 3D world points recorded during alignments into a 2D pixel equivalent. The difference between the reprojected point and the actual screen point used for the user alignments provides the *Reprojection Error* metric, Figure 3.9. Though this error measure is natively represented by a difference in pixels, a conversion to visual angle is possible using known FOV and resolution values for a specific display.

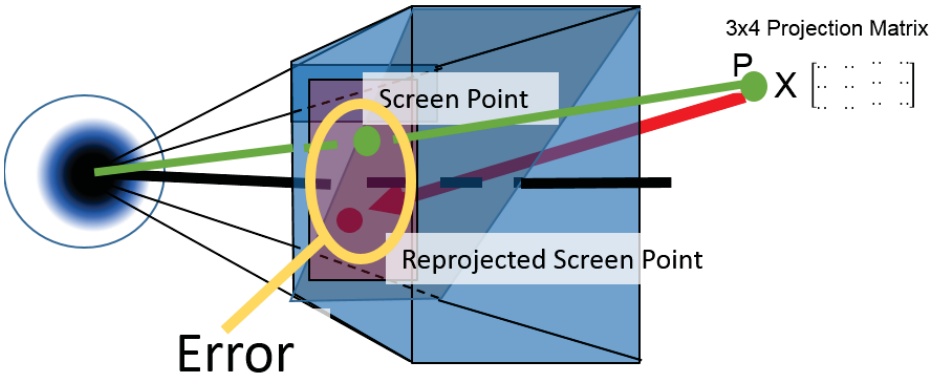


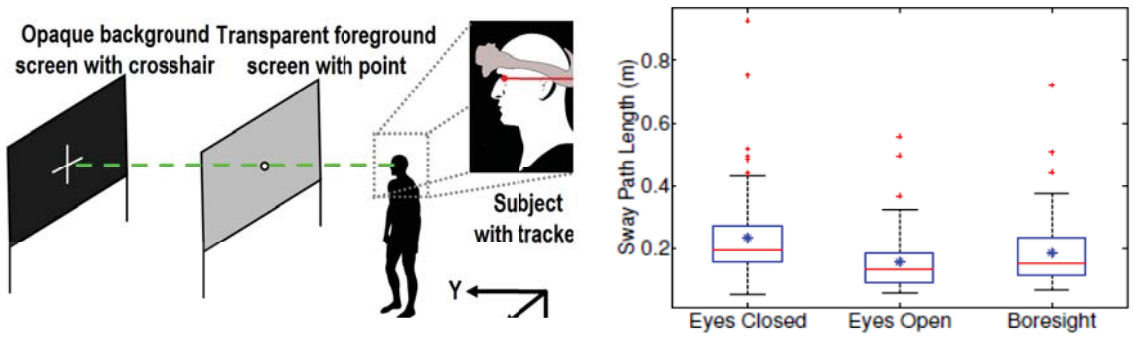
Figure 3.9 Illustration of reprojection error

The world point, P, used during the alignment phase is transformed using the 3x4 projection matrix produced by the calibration. The coordinates of the reprojected screen point are compared against the location of the on-screen point aligned with world point P

3.3.2 Evaluation Studies

While objective measures provide a hard quantifiable metric for evaluating a calibration result, it is often necessary to perform subjective studies as well to investigate the impact of various systemic and human factors on the perceptual quality of an AR system's registration. Studies by Axholt et al. [4, 5, 6, 7, 8, 9, 10], for example specifically investigate the issue of user misalignment during the correspondence phase of SPAAM calibration. Their studies focus, particularly, on the factors of motor control and postural sway which inhibit

and diminish the precision with which a person is able to perform a stationary alignment. The experimental setup utilized in studies [4, 5] is illustrated in Figure 3.10 (a). During this investigation, the postural stability, with regard to head motion, of participants was examined and Figure 3.10 (b) shows that the amount of sway is highly dependent on the visual load of the subject, with the most sway occurring while the users' eyes are closed.

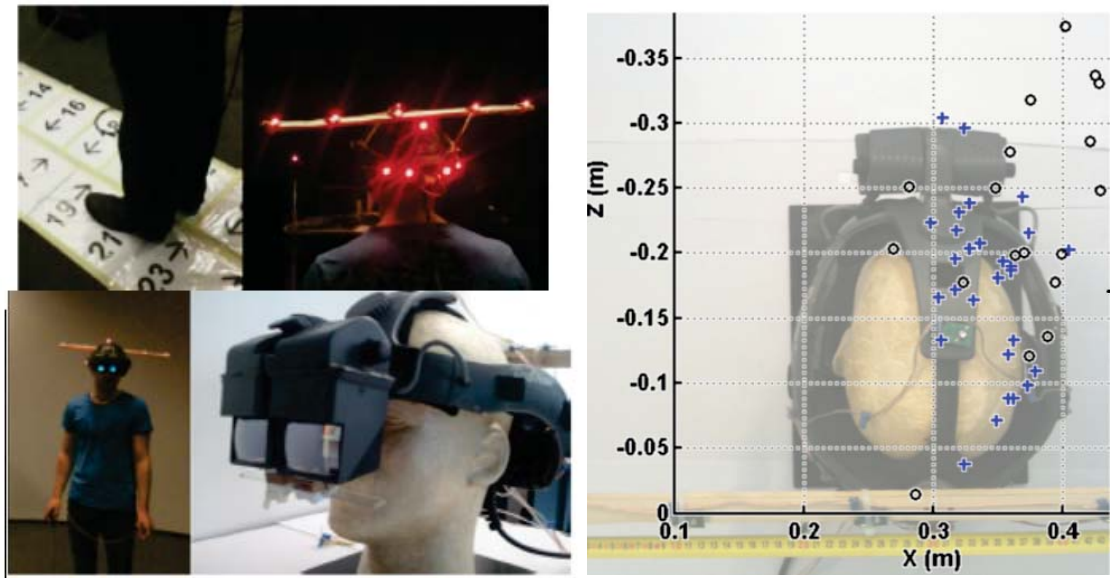


(a) (b)

Figure 3.10 Impact of visual load on postural sway

(a) Experimental setup illustrated from [4]
(b) Effect of sway length as a product of user visual load

In an attempt to ameliorate the impact of alignment inaccuracies, Axholt et al. [9, 10] further investigated the Depth SPAAM modality, arranging alignment distances more uniformly over the environment. This "Magic" distribution, so named for the use of a magic square to generate the distance intervals, showed significant improvement, with regard to objective extrinsic eye location estimates. Figure 3.11 shows the experimental setup and corresponding eye location estimates for the magic alignment distribution. A particular result of note is the larger variance in eye locations estimated laterally, in depth relative to the screen, compared to the horizontal and vertical axis.



(a) (b)

Figure 3.11 Investigation on alignment distance impact on SPAAM calibration

(a) Experimental setup from study [10]
(b) Extrinsic eye location estimates from study [10]. Blue crosses represent eye position estimates for calibrations using the "Magic" distance distribution. Circles represent results from alignments performed at sequentially changing distances

A related study by Maier et al. [84], examines the contribution that confirmation methods, for recording a user's alignment response, have on contributing error to the calibration. They consider standard entry mechanisms, such as keyboard and mouse, but also vocal response and timed input. Their results indicate that the timed input method, having the user hold the alignment for a set interval, resulted in more accurate calibration results over traditional input methods.

Perceptual evaluation studies, seeking to obtain information about perceived registration quality, will often utilize simple tasks through which an implicit metric of calibration accuracy can be obtained. Studies from Mcgarrity and Tang [87, 142] provide interaction methods for users to directly indicate the perceived registration of on-screen items using a

stylus and tablet. During these tasks, a virtual object is shown on-screen and the user uses the stylus to indicate the perceived 3D coordinate within the world where the virtual item appears to be registered. Navab et al. [104] extend the functionality of this approach by allowing users to also correct registration, through tangential and rotational shifts, during run-time. Grubert et al. [48] similarly conducted a user evaluation study of SPAAM and several variants, in which subjects indicated the real world correspondence point of on-screen items using a laser pointer. While providing a larger range or coverage compared to the sylus and tablet schemes, their discussion indicates that this method was quite time consuming for subjects to complete however, making it impractical for recording a large amount of registration data.

As noted in section 3.2.3, current advances in low cost miniature consumer devices holds much potential for advancing the usability and reliability of calibration procedures. Similarly, the current boom in the development of consumer level OST HMD devices is being met with a growing need for standardized ubiquitous calibration practices that are suited for use by inexperienced novice users. Of course, the growth in innovative calibration solutions must also be met with an equal rise in endeavors to generate equally novel evaluation processes.

CHAPTER 4

CONTRIBUTIONS

The novel work presented in this dissertation aims to enhance the body of knowledge pertaining to OST HMD calibration through updated user study evaluations focusing on the performance of not only state of the art calibration alternatives but also revised versions of standard approaches offering more versatile application to current consumer level head-mounted devices. A primary goal, therefore, is the development and evaluation of a user-centric approach adaptable to current and next generation OST hardware. Likewise, an investigation and comparison of the expected accuracy trade-offs for such a method, in comparison to the standard environment-centric implementations employed in prior studies, is an additional goal.

An area of particular intrigue is the correlation, if any, between the extrinsic eye location estimates from studies such as Axholt's and those produced by a user-centric calibration system. Contributions showcase the benefits of adopting user-centric calibration methodologies over traditional environment-centric schemes through the development and deployment of an actual environment-agnostic setup made possible by leveraging existing low cost consumer interface hardware. Additional work conceptualizing a novel method for on-line evaluation of calibration results by a third party observer is also presented along

with an improved strategy for enhancing the intuitiveness of stereo alignments for SPAAM calibrations targeting binocular HMDs. However, the production of a more in-depth comparison of semi-automatic calibration performance in contrast to the more common manual SPAAM methods is the first objective confronted in this work.

4.1 Study 1: Evaluation of Automatic vs Manual Calibration Methods

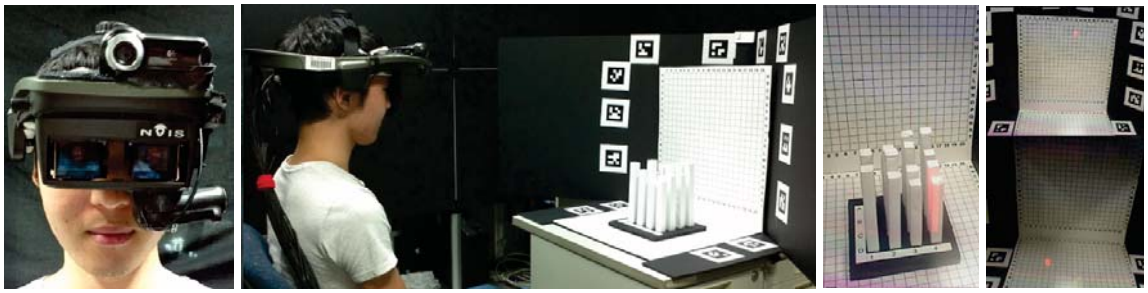
Motivated by the potential for automatic calibration solutions, this study formally evaluates Itoh and Klinker's INDICA calibration methodology against a traditional SPAAM implementation [93]. This is the first investigation to examine INDICA through a user study assessment employing both objective numeric metrics and subjective qualitative measures obtained through analysis of user performance in a registration critical task. The novelty of the experiment is further enhanced by the inclusion of a third calibration scenario, a degraded SPAAM condition, which commonly occurs in OST AR systems that reuse previous calibration results between uses without any subsequent update or recalibration to account for changes in HMD placement. Results from this experiment provide a base reference of expected performance–implementation trade-offs for each of the three calibration strategies, which is of especial importance to researchers requiring guidance to select the best calibration plan able to suit the complexity and accuracy constraints of their particular endeavor. Though a description of the experimental design, procedure, and results follow, the complete published work is available in [93].

4.1.1 Experimental Design

The construction and implementation of the investigation is conducted so that performance metrics for each of the three calibration conditions, INDICA, SPAAM, and Degraded SPAAM, are obtained after completion of the respective procedure. Further inspection of each of the three techniques is facilitated by user performance data obtained through two registration dependent tasks. A within-subjects strategy produces a total of six experimental conditions, 3 calibrations x 2 tasks, per subject. A total of 13 subjects, 6 male and 7 female, ultimately participate in the study, all of which possess normal or corrected-to-normal vision and have no prior experience using HMDs or OST AR applications.

The OST HMD system used during the study is composed of an NVIS ST50 binocular display with a resolution of 1280×1024, 40^o horizontal and 32^o vertical field of view, and spatial resolution of 1.88 arcmn/pxl. Even though the ST50 supports stereoscopic viewing, the right eye piece was purposefully obstructed to create a monocular viewing system. Limiting the view to a single eye not only simplifies the calibration procedures, but also prevents any inherent bias or performance issues that may arise from unknown stereo blindness or depth perception limitations within the subject pool. The 6 DOF pose of the HMD is determined through visible fiducial marker tracking facilitated by the Ubitrack software library [62] and a Logitech Quickcam Pro 9000 CMOS camera rigidly affixed to the anterior region of the display. The coordinate frame of the HMD is, likewise, defined by the tracking camera, with the origin located at the camera's viewing center. Remaining hardware consists of a second Logitech Quickcam, identical to the first, mounted below the left eye piece of the HMD. The sole use of this camera is to capture images of the user's

eye necessary for performing the iris detection and localization procedures for INDICA calibration. Figure 4.1 (a) provides a photograph of the complete HMD configuration and the location of the fiducial tracking and eye imaging cameras relative to the user's view.



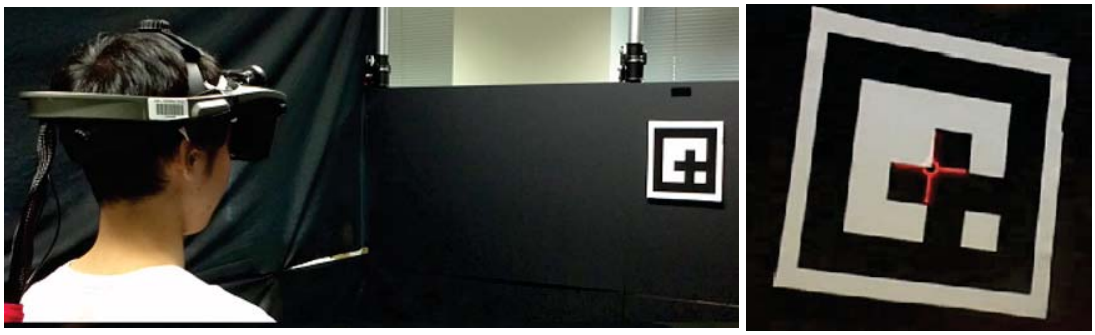
(a) (b) (c) (d)

Figure 4.1 Study 1 experiment setup and task design

(a) View of the HMD and attached marker and eye tracking camera
(b) Illustration of the location of the subjects relative to the task setup
(c) View of a virtual pillar as seen by the subjects
(d) View of a virtual cube on the vertical and horizontal cube grids

The Single Point Active Alignment Method as described in [148] is used as the control condition for the experiment. A total of 20 screen-to-world alignments is used to produce the calibration result, with each alignment performed by subjects visually aligning the center of an on-screen cross-hair with the center of a fiducial marker rigidly mounted within the world in front of them, Figure 4.2. The position of the fiducial marker, relative to the tracking camera mounted on the HMD, is tracked in real time and used, along with the 2D pixel coordinate of the on-screen crosshair, as the correspondence point for the Singular Value Decomposition (SVD) calculations used to generate the final projection matrix. The 2D pixel coordinates of each on-screen crosshair are chosen randomly at run time, and subjects are given the option to skip cross-hairs whose locations on screen make them

difficult to see. In order to reduce error due to subject movement during the alignment steps, a hand clicker is provided to subjects allowing them to non-verbally indicate when an adequate screen to world alignment is achieved. Subjects activate the clicker using one or more fingers, at which point the experimenter counts backward from 3 to 0 and records the correspondence measurement. During the calibration procedure, subjects are instructed to take a number of steps forward or backward so that alignments are performed at varying distances between 1.5m to 3m from the fiducial marker. Subjects only perform the SPAAM calibration once and always at the beginning of the experiment before any tasks are started.



(a) (b)

Figure 4.2 View of the SPAAM alignment process

(a) Location of the alignment target relative to the subjects
(b) View through the HMD of a correct screen–world alignment

The degraded SPAAM (DSPAAM) condition reuses the projection matrix produced by the subject's SPAAM calibration [64, 65]. This calibration method is chosen to replicate the real world condition where an HMD may shift or slip on a user's head, degrading the effectiveness of the calibration. To implement this condition, the HMD is simply removed from the subject and then replaced with only minimal care to ensure the subject's left eye

is within the exit pupil of the HMD and that on-screen visuals can be clearly seen. No further procedures are performed to correct any misalignment resulting from placement of the device.

The Recycled INDICA setup, described in detail in [64, 65], comprises the third calibration condition examined in this study. The Recycled INDICA variant generates a calibrated projection matrix by combining the intrinsic parameters obtained from decomposing the existing projection matrix produced by the subject's SPAAM calibration, with updated eye location extrinsics estimated at run-time. The extrinsic parameters are determined per the procedure outlined in [64], in which multiple images of the eye are taken and processed to identify the ellipse of the iris. The center and viewing direction of the eye is then approximated by projecting the ellipse into a spheroid in 3D space [140, 107]. This procedure is repeated over a sequence of 10 images, after which the median values for rotation and translation are combined with the existing intrinsic values to generate a final calibrated projection. Even though it is possible that the HMD position may shift during the experiment, the extrinsic eye locations are not updated once the tasks begin.

4.1.2 Tasks and Procedure

The objective measures discussed in section 3.3.1, those of eye location and reprojection error, are taken directly from the numeric calibration results. Additional subjective measures, intended to provide qualitative metrics for the accuracy of the registration produced by each calibration condition, are obtained through user responses during two similar but distinct visual tasks. Both charge participants with determining the 3D location of

a virtual object. However, the range of allowable responses is limited to a pre-defined set of discrete values. In addition to the perceived registration location, a second independent measure describing the quality of the registration, how well the virtual object is overlaid to the chosen location, is also recorded.

4.1.2.1 Pillars

Participants are tasked with indicating which, out of 16, real world pillars an on-screen virtual pillar appears to be best registered with. The virtual pillar is rendered at each of the real pillar locations once, for a total of 16 trials per calibration method. Figure 4.1 (c) shows the real pillar arrangement with on-screen virtual pillar, rendered in red, as it would appear during the task. During each measurement, the subject is able to freely choose any one of the sixteen real pillars, denoted by a letter and number combination according to the row and column ordering, they feel the virtual pillar is best aligned to. The ordering of virtual pillar locations is randomly permuted under the constraint that the next pillar location is chosen to be in both a different row and column as the previous. Heights for the real pillars cover the range 13.5 cm–19.5 cm varying by .25 cm increments. The pillars are arranged in a 4×4 grid such that the average height of the pillars in each row and column is between 16.25 cm–16.75 cm. The virtual pillar, displayed on-screen, is rendered such that it should appear to be a constant height of 15.5 cm. Once the virtual pillar is displayed at all sixteen real pillar locations, the task ends.

Subjects also verbally provide a quality rating for each trial of the task. A 1 to 5 subjective scale, with 1 denoting the worst registration and 5 denoting the best registration,

are used for this metric. Before beginning the task, subjects are informed of the quality scale and provided printed images illustrating the expected visual quality that should be present at each quality level. The top row of Figure 4.3 shows the quality scale reference images provided to each user for the Pillars task.

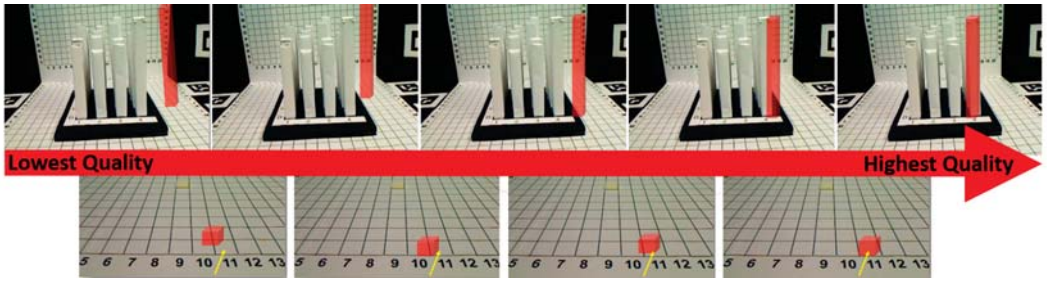


Figure 4.3 Quality scale images provided to subjects prior to performing each task

Each view represents the approximate registration required for each level with quality increasing left to right along the scale. Top row for Pillars task quality, bottom row for Cubes task quality

4.1.2.2 Cubes

Participants are tasked with indicating which, out of a possible 400, grid locations a virtual cube appears to be best registered with. Two separate grids are used for this task, each comprised of $2cm \times 2cm$ squares in a 20×20 arrangement. Rows for each grid are labeled with letters from A-T and columns labeled with numbers from 1–20. The first grid is positioned flat on the task table in front of the user and is referred to as the horizontal cubes grid. The second grid is placed perpendicular to the horizontal cubes grid so that it faces the user. This perpendicular grid is referred to as the vertical cubes grid. The complete arrangement used for the task can be seen in Figure 4.1 (d). The virtual cube, shown on the HMD, is modeled such that its perceived size should be approximately

$2cm \times 2cm \times 2cm$ and rendered red for increased contrast with the real environment. The virtual cube is presented at 10 grid locations on both the horizontal and vertical grid for a total of 20 trials per calibration condition. The positions of the virtual cube, on either grid, are randomly selected such that no location is repeated. The display order is chosen such that no consecutive virtual cubes will appear in the same row or column. Ordering of trials between the horizontal and vertical cubes grid locations are also selected randomly, and subjects are verbally informed at the start of each trial which grid the virtual cube should appear upon. For each of the 20 trials, subjects indicate their selection by stating the row letter followed by the column number of the grid location to which they feel the virtual cube is best aligned.

Subjects also verbally provide a quality value for each trial of the task. A 1 to 4 subjective scale, with 1 denoting the worst registration and 4 denoting the best registration, are used for this metric. Before beginning the task, subjects are informed of the quality scale and provided images illustrating the expected visual quality that should be present at each quality level. The bottom row of Figure 4.3 shows the quality scale reference images provided to each user for the cubes task.

4.1.3 Study Results

4.1.3.1 Objective Measures

The two objective metrics considered in this study are the extrinsic eye position estimates and reprojection error determined from the screen to world correspondence pairs recorded during the SPAAM calibration. The degraded SPAAM condition, though, is not

considered for the quantitative analysis since an identical projection matrix is used for both this and the SPAAM condition. Figure 4.4 provides a comparison of eye position estimates for both SPAAM and Recycled INDICA. The plots show the mean and variance of values across all 13 subjects. All axis positions are relative to the display screen, with X along the horizontal and Y along the vertical screen direction. The Z axis is distance from the display screen toward the user. Both SPAAM and Recycled INDICA produce similar eye position estimates. Not surprisingly, values along the Z direction are less varied, similar across all subjects, using the Recycled INDICA eye imaging method. The large Z axis variance for the SPAAM extrinsic estimates though, corresponds to a similar pattern found by Axholt et al. [10]. Differences in estimates along the Y and X axis are substantially more similar for both conditions, with the SPAAM Y axis positions being slightly more consistent than those recorded through the INDICA methodology.

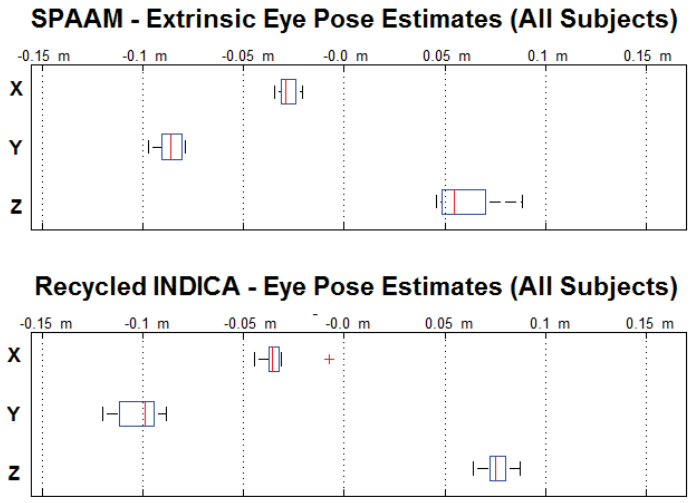
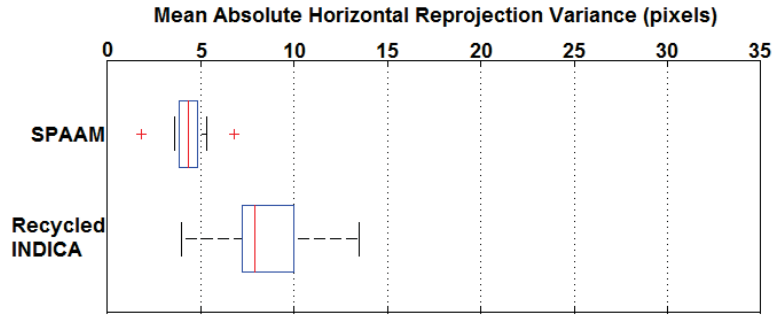


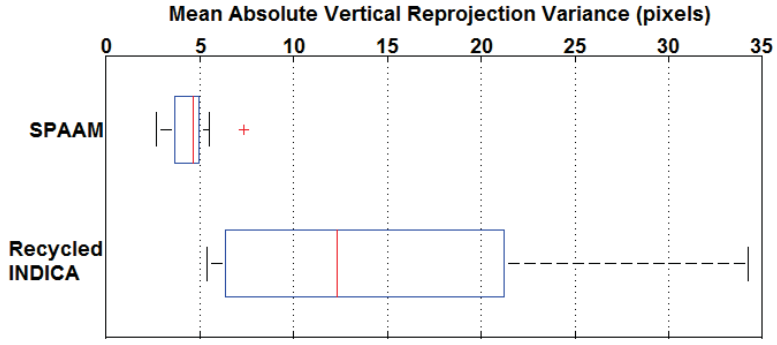
Figure 4.4 Eye position estimates across subjects for SPAAM and Recycled INDICA

Axis are relative to the display screen, with X along the horizontal and Y along the vertical screen direction. Positive Z is away from the display screen toward the user. All values are in meters

Since the Recycled INDICA condition does not require the use of screen to world alignments, the 2D–3D point correspondence pairs recorded during the SPAAM calibration are used to produce reprojection error values for both the SPAAM and INDICA conditions. This error is calculated as the difference in pixel location between the result of reprojection, transforming the 3D world point into screen space using the projection matrix result, and the actual 2D screen location of the croshair used during the SPAAM alignment. Figure 4.5 provides the error, in terms of pixel differences, for all subjects along the horizontal and vertical screen axis respectively.



(a)



(b)

Figure 4.5 Absolute reprojection variance for SPAAM and Recycled INDICA

(**Top**) Absolute reprojection variance in horizontal screen space for SPAAM and INDICA (**Bottom**) Absolute reprojection variance in vertical screen space for SPAAM and INDICA

4.1.3.2 Subjective Measures

The two subjective metrics considered in this study are the perceived location of the virtual object and the quality of the registration at that location. Error in perceived location is taken as the difference between the subject reported row/column position and the actual intended location where the virtual object should have appeared. The difference along a row indicates registration error in the horizontal, X, direction relative to the tracking coordinate frame, with negative error indicating a user value that is to the left of the intended position. The difference along a column represents error in the vertical, Y, direction for measures taken during a trial on the vertical cubes grid, with negative error indicating a user value that is below the intended position. Difference along a column in both the pillars and horizontal cube grid trials is interpreted as error in distance, Z, relative to the tracking coordinate frame, with negative error indicating a response that is closer to the user than the intended position. A conversion of the error measure is also performed to interpret the difference in grid squares to distance measures. The size of grid squares for the cubes task is $2cm \times 2cm$. Thus, we equate an error of 1 square in any direction to an error of 2cm in the respective direction. Similarly, the spacing of pillars in the pillars task is 4cm, since each $2cm \times 2cm$ pillar is separated by a 2cm row or column. Therefore, an error of 1 pillar is equated to an error of 4cm in the respective direction. A reduced nomenclature is also adopted for presenting the results for the cubes tasks, Cubes-V representing measures for the vertical cubes grid and Cubes-H representing measures for the horizontal cubes grid.

Figure 4.6 provides the distance converted registration error results for the Pillars task. Error in both the X, Left-Right, and Z, Front-Back, directions relative to the tracking coor-

dinate frame are provided. Repeated-measures analysis of variance (ANOVA) performed across the X dimension error shows no significant main effect due to calibration method $(F < 1)$, with each condition producing nearly perfect, 0 error. All three calibration methods, however, do produce error in the Z direction, with subjects perceiving the registration of virtual objects to be closer than intended for every case, with ANOVA revealing a highly significant effect of calibration method $(F(2, 24) = 14.011, p < 0.001)$. Recycled INDICA, though, produces a shift in perceived distance closer to the correct location, compared to the other conditions.

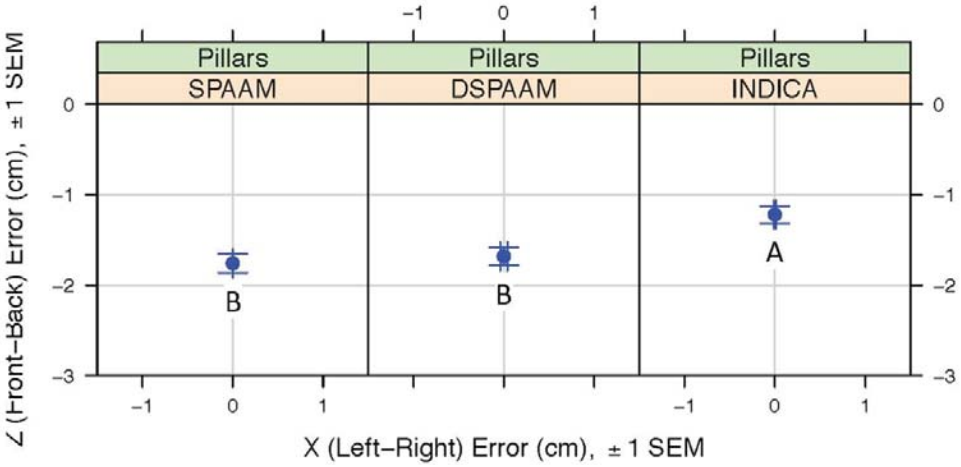
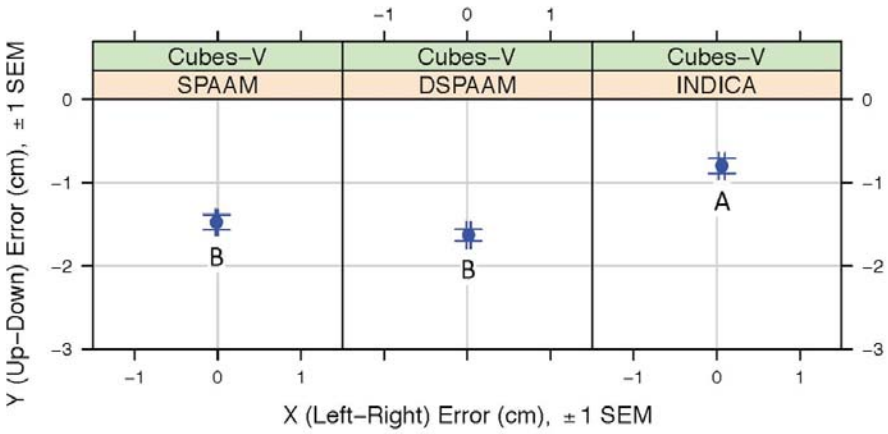


Figure 4.6 Pillars task grid error along the X (Left-Right) and Z (Front-Back) axis

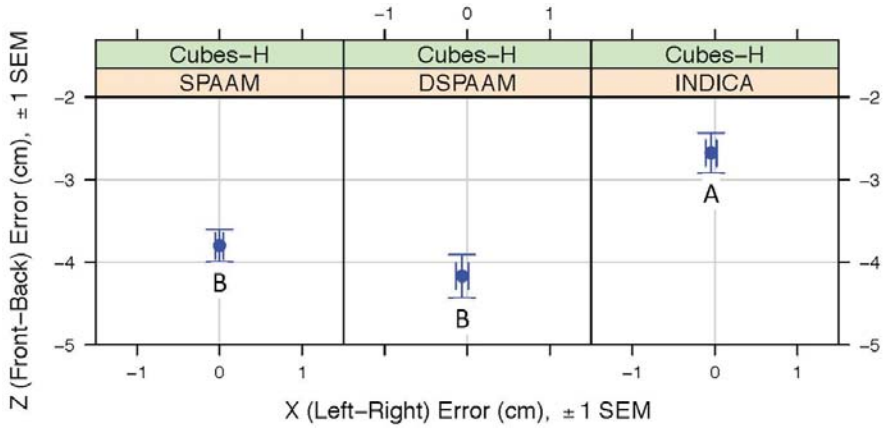
Pillars task error along the X (Left-Right) and Z (Front-Back) axis

Figures 4.7 (a) and (b) show the distance converted registration error results for the Cubes task separated by each grid, Cubes-V and Cubes-H respectively. ANOVA performed across Cubes-H measures shows a significant main effect of calibration method along the Z direction $(F(2, 24) = 7.37, p = 0.003)$, and no effect along the X $(F < 1)$. Similar to the Pillar task results, all three calibration methods produce equally near 0 error along the

X direction with Recycled- INDICA produces the lowest error in the Z direction. Cubes-V measures shows nearly identical results, no main effect along X $(F < 1)$, and Recycled INDICA producing a positive effect along Y $(F(2, 24) = 10.96, p = 0.0016, e = 0.75)$.



(a)



(b)

Figure 4.7 Task grid errors

(a) Vertical cubes grid task error along the Y (Up-Down) and X (Left-Right) axis
(b) Horizontal cubes grid task error along the Z (Front-Back) and X (Left-Right) axis

The subject-provided quality, shown in Figure 4.8, are also normalized for analysis, since the respective scales differ for each task. Measures for both tasks are normalized to values from 1 to 4, which does not change the user specified values for the cube task but does compress the scale for quality values recorded for the pillars. Converting both tasks to an identical scale allows for direct and fair comparisons between tasks across subjects. A significant main effect of calibration method occurs in both the pillars task $(F(2,24) = 5.03, p = 0.015)$ and the Cubes-H grid $(F(2,24) = 6.65, p = 0.013, e = 0.71)$. The Cubes-V grid condition shows no significant difference between calibration method $(F < 1)$. The plots of Figure 4.8 also reveals that subjects perceived Recycled INDICA registrations, viewed on the pillars and the Cubes-H grid, to be of higher quality over Degraded SPAAM. Also, while SPAAM quality is rated nearly equal to Recycled INDICA in the pillars task, it rates lowest in Cubes-H grid trials overall. All three calibration methods produce nearly identical quality ratings across subjects in Cubes-V trials.

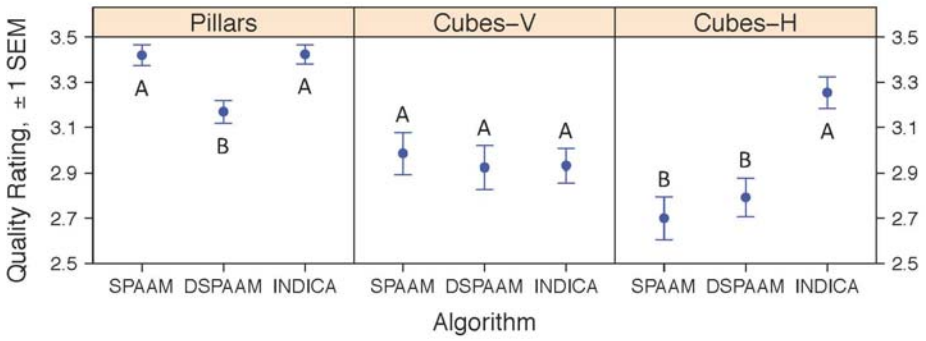


Figure 4.8 Mean subjective quality values for each calibration method during each task Values normalized to a 1–4 scale with 1 denoting the lowest quality and 4 the highest

4.1.4 Discussion and Conclusion

As noted earlier, the eye position estimates for SPAAM, Figure 4.4, show larger variance relative to frontal screen distance, which matches closely with similar findings from previous studies [10, 8, 64]. Consequently, the eye estimates along Z for Recycled INDICA show much smaller variance compared to the other conditions. While the variability in the findings do differ from Itoh and Klinker's initial results, this difference is undoubtedly a byproduct of the multiple user study design, whereas Itoh and Klinker's results derive from a single user. The reprojection estimates from this study also differ from those presented in [64], which indicate that Recycled INDICA should produce errors with similar variance to SPAAM. Figure 4.5, however, shows that Recycled INDICA reprojection error is significantly higher, particularly in vertical screen space. It is reasonable to conclude that the SPAAM performance of the subjects contributed to this disparity in reported findings. The correspondence pairs used for the reprojection error calculation presume a perfect alignment was created between the center of the crosshair and the recorded 3D point location. Since all subjects were completely unfamiliar with the alignment procedure, and HMD's in general, the precision with which alignments were performed will greatly vary. The SPAAM solution itself is tailored to minimize error by fitting the solution to best match the correspondence point pairs. Therefore, it is logical that the reprojection error will be lowest for SPAAM, and will result in higher error for INDICA. The high reprojection error, in this case, does not reflect a negative performance of INDICA, but instead provides an indication of the actual alignment error incurred by subjects during the SPAAM procedure. Even though alignment accuracy may have directly impacted the quality of the SPAAM

calibration, the subjective registration accuracy measures show only a slight deviation in performance between conditions.

According to Figures 4.6 and 4.7, all three calibration techniques produce virtual content registration that is perceived as being closer to the subject than intended. Since users were restricted to viewing images through only the left eye piece, it is possible that the lack of stereo depth cues influenced this underestimation of registration location. The larger error in the Z location for the extrinsic eye estimates in the SPAAM condition is also a likely factor for this result as well, given that the updated eye estimates for Recycled INDICA appear to have had a correcting effect on the perceived registration distance, as well as in the vertical field of view. It is also interesting to note that all three calibration techniques produce nearly perfect registration in the horizontal direction. It is yet unclear whether this correlates to the similar eye location estimates in the X direction seen in Figure 4.7, or because the object position in the X direction is easier to isolate due to the availability of multiple viewing angles, from subjects leaning forward, backward, and sideways during the tasks.

An additional item of note is the difference in significance produced by the ANOVA analysis between the perceived quality and registration results. The quality values for trials on the vertical cubes grid show no significant difference even though the error measures along the Y direction, show significance. A similar result can be seen for quality values on the horizontal cubes grid and error measures along the X direction. This discrepancy may be partially due to the inclusion of both directions for the quality values, whereas the error plots show results for each direction in isolation. The experimental design did not facili-

tate the recording of independent qualities for each direction of the grid, and, therefore, it must be inferred that the quality evaluations are based on the perceived registration along the X and Y direction together. The analysis does clearly show, however, that subjects felt the overall quality of the Recycled INDICA registrations to be higher in comparison to SPAAM and Degraded SPAAM. It can be safely presumed that the higher subjective quality given to Recycled INDICA directly correlates to the higher registration accuracy observed in the tasks. This implies that non-expert users rely heavily on perceived registration location for information, an important item of consideration for AR designers.

This experimental study has shown that the Recycled INDICA OST HMD calibration method has the potential to produce registration that is both more accurate and of subjectively higher quality than the common SPAAM based calibration techniques, especially in regards to registration perceived in depth. It can be further noted that the performance of Recycled INDICA will degrade far slower than that of interaction dependent methods, due to the unreliance on correspondence alignments.

A drawback to implementing INDICA, though, is the need for eye imaging hardware. Nearly all of the currently available OST HMD's are not factory equipped with the required eye tracking cameras, and thus it is up to the investigator to suitably mount the necessary equipment. However, this study also shows that a degraded SPAAM condition, in which calibration results are reused without updates to display position, does not produce any significant degradation in perceived accuracy or registration quality. This finding has important implications for those individuals desiring to use OST AR for applications where

recalibration time needs to be minimized and a minor level of registration inaccuracy is acceptable.

Though the INDICA results were more favorable than those for SPAAM, a simplified process, able to ameliorate errors incurred during the alignment process, may not only improve the perceived quality and accuracy of the SPAAM calibration, but would also improve the accessibility of the method even more for current generation HMD hardware, compared to the implementation costs of an INDICA approach.

4.2 Study 2: Evaluation of User-Centric SPAAM Calibration using Leap Motion

Though the results of Study 1 showed that the automatic INDICA calibration has the potential to provide greater accuracy and perceived quality of virtual content registration, the requirement of additional eye tracking hardware and algorithms makes this methodology largely inaccessible for application to present OST display offerings. With manual user dependent approaches, such as SPAAM, remaining the only viable calibration option, motivation for this second study arose from the need to shift research focus toward the development and evaluation of easily standardized procedures with low implementation and user performance costs that are also appropriate for use with current and next generation hardware.

This experiment is based around a two-fold objective. The first goal is to provide an evaluation of a SPAAM implementation that does not rely on any rigid environment features. This purely user-centric approach must, therefore, be constructed with the intent to use trackable features of the user's person instead of fiducial markers or other pre-measured

locations within the tracking space. The second goal of the study is to devise a low-cost implementation strategy, for this environment-agnostic calibration, that leverages existing and readily accessible consumer hardware.

Both of these objectives are accomplished in this work, investigating the efficacy of the Leap Motion controller as a means for facilitating hand and stylus alignment based calibration of an OST HMD system. In contrast to a previous cursory demonstration [67], this study additionally includes an examination of accuracy and precision differences between monocular and stereo calibration variants. Also explored are several reticle designs and the effect of alignment context on hand calibration results. The analysis employs standard objective measures, including the reprojection error and extrinsic eye location metrics used in Study 1, to compare not only the performance of each condition, but also the viability of OST calibration with Leap Motion in general. The outcomes of the study directly benefit efforts toward devising standardized calibration practices, and provide much needed insight into the viability of SPAAM like calibration approaches unreliant on an established tracking or environment frame of reference. Though a description of the experimental design, procedure, and results follow, the complete published work is available in [98, 99].

4.2.1 Experimental Design

A complete OST AR framework is constructed by combining a Leap Motion controller with a commercially available HMD. Since the Leap Motion is able to perform both hand and stylus tracking, both mechanisms are utilized for performing monocular and stereo SPAAM based calibrations, requiring multiple alignments between on-screen reticles and

either a finger or stylus. While achieving consistent alignments to a single point on a stylus is relatively intuitive, the ability of a user to maintain repeated alignments with a single point on a finger tip is far more inexact. Instead of utilizing an additional physical cap, ring, or other wearable indicator, variations of the on-screen reticle design were created to provide visual context for aiding the user in properly positioning their finger during calibration alignments. This approach was taken to more adequately represent what a viable consumer oriented calibration mechanism would provide.

The same NVIS ST50 binocular OST HMD, used for Study 1, is also used as the primary display for this investigation. Unlike Study 1, however, the binocular capabilities of the display were fully utilized for stereo calibration. A custom 3D printed mount is created to attach the Leap Motion to the anterior of the display. Figure 4.9 (a) and (b) show the complete assembly and orientation of the Leap Motion tracking coordinate frame relative to the HMD. Integration of the Leap Motion tracking information is performed using version 2.3.1.31549 of the available SDK. The remaining piece of external hardware created for the system is the stylus rod. A stylus, or tool as it is referred to in the Leap Motion documentation, refers to any simple cylindrical object of sufficient length and diameter to be seen and recognized by the device. The stylus tool used in this study is created from a 5mm diameter wooden dowel rod, approximately 20cm in length to allow the majority of the rod to extend beyond a user's hand when held.

(a) (b) (c) (d)

Figure 4.9 Study 2 hardware setup

(a) The right-handed coordinate frame of the Leap Motion
(b) Combined HMD and Leap Motion apparatus
(c) User performing a screen to stylus alignment
(d) Rendered skeleton overlayed onto the user's hand

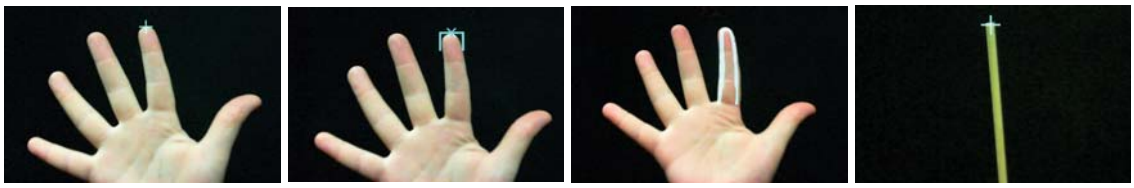
4.2.2 Alignment Methods and Procedures

4.2.2.1 Hand Alignments

Tracking data from the Leap Motion is able to provide the position and orientation of numerous points along the hands and arms, as long as they are within the field of view of the device. Alignment complexity for the SPAAM procedure is reduced by restricting the correspondence to the position of the tip of the right index finger. Even though a specific finger is used, defining the exact location of the tip is far more ambiguous. In order to provide greater context to the user during the alignment phase of the calibration, three separate on-screen reticle designs were employed. A perfect alignment occurs when the center of the right index finger tip coincides with the target point specified for each reticle.

The first, and most generic design, is a simple cross-hair comprised of a horizontal and vertical line, displayed with the target point located at the center of the intersection point. The on-screen dimensions of the cross are 64×64 pixels with line thickness of 3 pixels.

The ubiquitous application of cross-hairs for targeting and aiming purposes makes this a natural design for alignment procedures. Figure 4.10 (a) illustrates an alignment between a hand and cross reticle as viewed through the HMD system.



(a) (b) (c) (d)

Figure 4.10 Stylus and hand alignments for each reticle design, as seen through the HMD.

(a) Cross reticle alignment.
(b) Box reticle alignment.
(c) Finger reticle alignment.
(d) Stylus alignment.

The second reticle is crafted to mimic a cap for the user's finger. This box reticle is created from a 3 sided rectangle displayed with an X placed on the upper edge intersecting the target alignment point. The onscreen dimensions of the box itself are 128×128 pixels with line and 'X' thickness of 10 and 5 pixels respectively. The structure of the box design is such that a user would naturally center their finger within the outlined region, improving the likely hood of consistent alignment to the finger tip. Figure 4.10 (b) illustrates an alignment between a hand and box reticle as viewed by a user through the HMD.

The final reticle provides an anatomical finger outline onto which the user's real finger is aligned. The target point for this reticle is located at the center tip of the outline's upper edge. The on-screen dimensions of the finger outline are 128×384 pixels with average thickness of 20 pixels. The finger reticle is intended to provide the most position context

of all three designs. The center portion of the reticle is not filled in order to allow the user a clear view of their finger throughout alignment. While it is possible to provide a completely solid design, the brightness of the display often inhibits the ability to clearly distinguish the location of real objects behind AR content. Figure 4.10 (c) illustrates an alignment between a hand and finger reticle as seen through the display.

4.2.2.2 Stylus Alignments

Tracking data from the Leap Motion provides not only the diameter and length of a tracked stylus tool, but also the tip position and pointing direction. Only the 3D tip position is considered for the calibration process. The same cross reticle used for finger alignments is also utilized for the stylus calibration condition. A perfect alignment occurs when the center of the cross coincides with the center of the stylus tip. Though additional reticle designs for stylus alignment could have been implemented, denoting the tip center for the stylus is extremely unambiguous and context, therefore, for this condition was not a factor of consideration. Figure 4.10 (d) illustrates a stylus alignment as viewed by a user through the HMD.

4.2.2.3 SPAAM Procedure

A standard alignment based SPAAM procedure is followed to calibrate the OST AR system. As recommended in Axholt et al. [10], a total of 25 alignments is used to generate the final calibration results. The on-screen points are distributed within a 5×5 grid pattern. Monocular, each eye sequentially, and binocular, both eyes simultaneously, calibration schemes are employed. Stereo calibration is facilitated by shifting the on-screen location of

all reticles for each eye to induce stereopsis, and the perception that the reticles are rendered at depth. While the placement of on-screen reticles differs between left and right eyes, no change is made for the patterns between monocular and stereo calibrations, or across calibration sets, to enforce consistent use of on-screen coverage regardless of condition

Both hand and stylus calibration conditions proceed in an identical manner. A single on-screen reticle is rendered to the display screen. The user then moves their right index finger or stylus tip until it aligns as closely as possible to the target point of the reticle, as previously described. Once a sufficiently accurate alignment has been achieved, a button press, on either a keyboard or wireless controller activates recording of positional data acquired by the Leap Motion. Throughout the recording process, the color of the on-screen reticle is changed from green to yellow providing visual confirmation to the user that measurement has begun and to indicate that the alignment should be maintained until recording has ceased. The 3D finger or stylus tip location, relative to the Leap Motion coordinate frame, is measured once every 100ms for 1sec resulting in 10 data points per alignment. The median X, Y, and Z position value from the 10 recording points is used as the final measure. This location estimate, along with the X, Y, screen pixel location of the reticle's target point, is saved, and the complete set of 25 world and screen correspondence pairs is combined to produce the calibration result.

Monocular calibration sets always proceed by first calibrating the left eye followed by the right without interruption. Stereo calibration sets, of course, produce both left and right results together. Additionally, the user is instructed to perform all hand alignments in an identical manner, by keeping their palm flat and facing toward the display screen with all

five fingers as evenly spaced as possible. This requirement is imposed to maintain a hand tracking quality as consistent as possible across conditions. All stylus alignments are also performed with the user holding the stylus in their right hand. No further restrictions are placed on the alignment procedure.

4.2.2.4 Participant

All calibration data is recorded from repeated trials by a single expert user. Since the primary objective of this study is to verify the efficacy of the Leap Motion controller itself, for calibrating OST displays, and not the inherent usability or intuitiveness of the design, repeated measures from an expert user, knowledgeable with the procedure as similarly employed by [65, 110], provide more stable results, void of subjective affects. The expert subject completed 20 monocular and 20 stereo calibrations for each of the three hand and single stylus alignment methods, resulting in 20 x 4 x 2 = 160 calibrations total. The user's maximum IPD is also measured to be approximately 62mm.

4.2.3 Study Results

The quality of the calibrations produced by each condition is evaluated using three primary metrics. The first two are identical to the objective measures from Study 1: estimated eye location of the user relative to the HMD coordinate frame obtained by decomposing the extrinsic component from the calibration results, and reprojection error calculated as the difference between the ground truth on-screen position of each reticle, used during calibration, and the screen coordinate that results from back projecting the corresponding finger or stylus 3D tip position using the projection matrix results. The third metric examines

binocular disparity values, taken as the difference between the left and right eye location estimates for each monocular and stereo calibration. This metric was not facilitated by Study 1 simply because calibration data was only available for a single eye. Study 2 explicitly examines the stereo calibration condition allowing this metric to be fully utilized.

4.2.3.1 Eye Location Estimates

Figure 4.11 provides the estimated eye locations obtained from the monocular and stereo calibration results of each alignment method. The 3D eye positions are provided in Figure 4.11 (a) through (d) for the Cross, Box, Finger, and Stylus alignment method results respectively. Likewise, a top-down view, showing only the positions relative to the X and Y axis of the HMD coordinate frame, are plotted for each alignment condition in Figure 4.11 (e) through (h). Visual inspection of the 3D plots show that the stylus alignment method produced the most accurate extrinsic results, in relation to plausible ground truth eye positions, green circles, over all. The highest level of precision, for both monocular and stereo calibrations, likewise occurs for stylus alignments. It can also be seen, through to a lesser extent, that the more contextual reticle styles increased the stability, clustering, of the extrinsic estimates. The 2D plots further reveal that the primary location error occurs along the depth dimension, the Y axis in this study, corresponding to similar results from SPAAM seen in Study 1 and those from Axholt et al. [10]. As in the 3D plots, the 2D cross-section shows that stylus alignments produced very little variation in depth, contrasting strikingly to the expected distributions found for all three finger alignment conditions. The increased

clustering within the contextual reticle conditions for finger alignments is present, though the level of improvement is significantly less compared to the stylus condition.

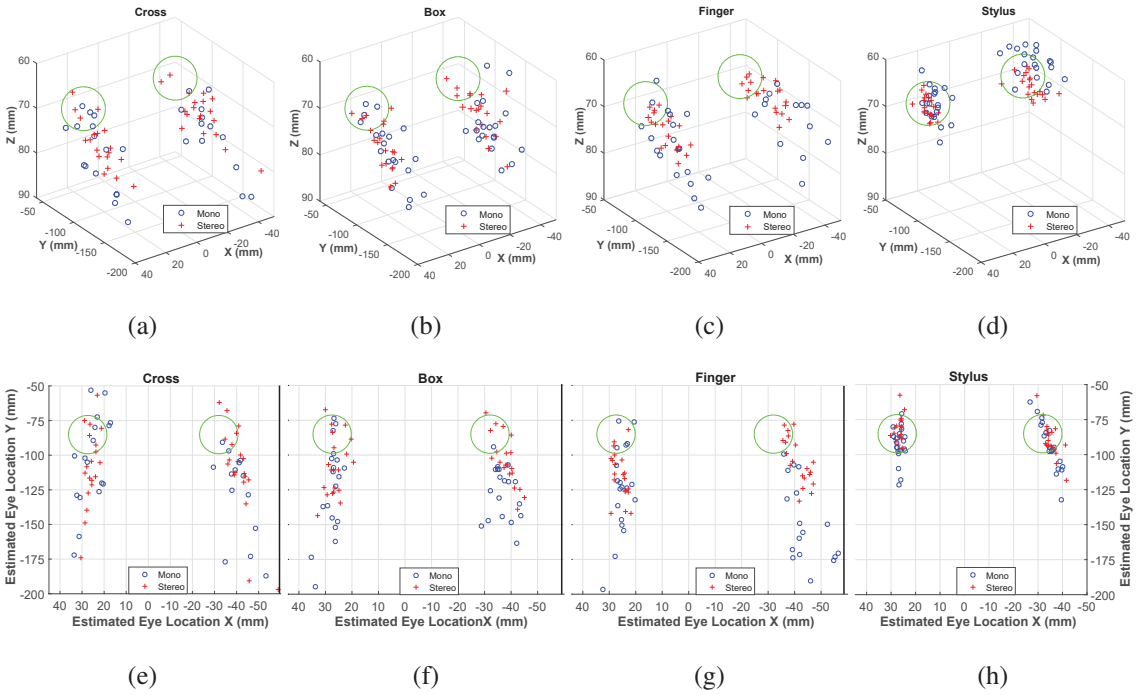
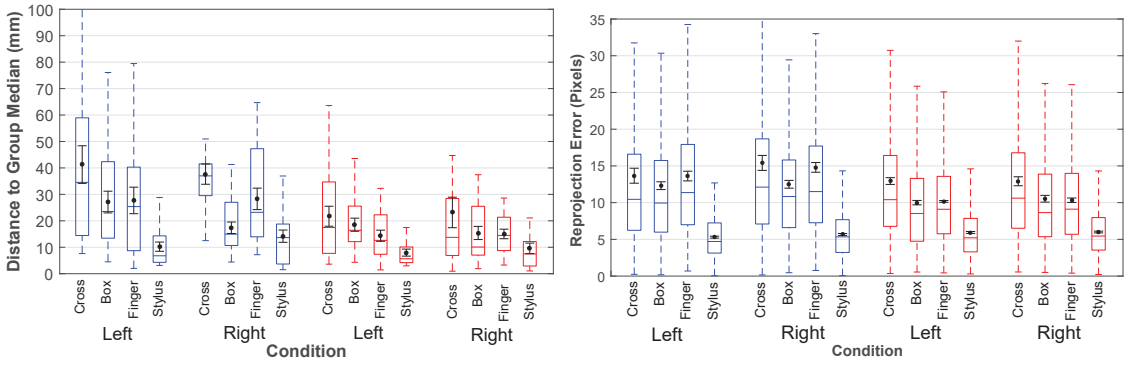


Figure 4.11 Estimated user eye locations relative to the Leap Motion coordinate frame

(a) Cross reticle, (b) Box reticle, (c) Finger reticle, and (d) Stylus calibration 3D position estimates. 2D eye position plots showing only X and Y estimate locations for (e) Cross reticle, (f) Box reticle, (g) Finger reticle, and (h) Stylus calibrations. In all plots, the center of the Leap Motion is at location (0, 0, 0), with monocular calibration estimates displayed in blue, stereo calibration estimates plotted in red, and green circles used to denote an area of plausible eye points

The amount of variance, or spread, in location estimates relative to the centroid value of each related group is also calculated and provided in Figure 4.12 (a). ANOVA between conditions identifies a significant difference between the four alignment method conditions in relation to both monocular and stereo calibration results. Significant differences are identified between alignment methods for monocular calibrations $(F(3, 57) = 13.3, p < 0.001)$, with a post-hoc Tukey-Kramer honest significant difference test confirming that

the cross alignment method median group distances are significantly higher compared to the remaining three methods at $(p < 0.001)$. Significant differences are similarly indicated between stereo calibration conditions $(F(3, 57) = 4.7, p = 0.01, = 0.68)$, with post-hoc analysis confirming that results from the cross alignment condition are significantly higher than the finger and stylus conditions at $(p < 0.001)$. Comparison of the monocular and stereo results within each alignment method also finds that the cross condition significantly differs between mono and stereo calibration $(F(1, 19) = 7.7, p = 0.01)$. Median distances for the finger alignment methods also differ significantly between monocular and stereo calibrations $(F(1, 19) = 15.6, p < 0.001)$.



(a) (b)

Figure 4.12 Mean distance and reprojection errors

(a) Distances between estimated eye positions and the median location value for monocular (blue) and stereo (red) calibrations for each alignment method
(b) Reprojection error for monocular (blue) and stereo (red) calibrations of each alignment method

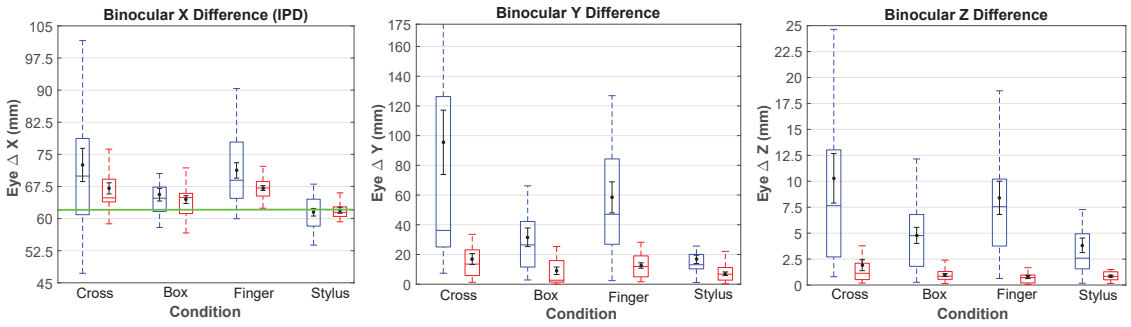
4.2.3.2 Reprojection Error

Figure 4.12 (b) provides the complete set of reprojection error values for both mono and stereo calibrations within each of the four alignment methods. Visual inspection and

ANOVA reveals a significant difference in calibration results across alignment method for both monocular, $(F(3, 117) = 41.0, p < 0.001, = 0.68)$ and stereo $(F(3, 117) = 45.8, p < 0.001, = 0.54)$ conditions. Post-hoc analysis confirms that reprojection error values for both the cross and stylus alignment method are significantly different from all other methods $(p < 0.001)$. Additional results between mono and stereo calibrations of each alignment method show only a significant difference errors within the box alignment method $(F(1, 39) = 8.7, p = 0.005)$, as well as for the finger and stylus methods $(F(1, 39) = 30.6, p < 0.001$ and $F(1, 39) = 9.4, p = 0.004$, respectively).

4.2.3.3 Binocular X, Y, Z Disparity

The three binocular disparity values, determined as the difference between the separate X, Y, and Z components of the paired left–right eye location estimates, are provided in Figure 4.13 (a), (b), and (c), respectively. While the IPD of the user is measured to be 62mm, the ground truth physical differences in the Y, depth, and Z, vertical, offsets are not directly determined, but are reasonably expected to be approximately 0mm. Visual inspection reveals that stereo methods, for each alignment condition, significantly out performs the monocular counterpart. Though, while IPD estimates for the stereo stylus condition match nearly perfectly the measured value of the user, the Y and Z disparity values show almost no improvement gains compared to the three finger alignment conditions. The reader is referred to the published work in [98, 99] for the complete exposition of ANOVA results.



(a) (b) (c)

Figure 4.13 Differences between left and right eye location estimates

Difference between X (a) positions represents interpupillary distance (IPD). The green line indicates the measured IPD, 62mm, of the expert user. Y (b) position differences indicate forward and backward offsets and Z (c) vertical offsets between left and right eye estimates in relation to the Leap Motion coordinate frame. Mean and ± 1 SEM bars are shown for the values within each condition group, (blue) for monocular and (red) for stereo

4.2.4 Discussion and Conclusion

Across all three error metrics, stereo calibration performs consistently better compared to the monocular variant. This can be observed visually by the tighter clustering of eye points in both the 3D and 2D plots of Figure 4.11 and across all three disparity measures of Figure 4.13 (a), (b), and (c). A clear distinction, though, is also present between the alignment methods themselves. Calibrations performed with the stylus produce lower distances to group medians and the highest level of accuracy in comparison to plausible eye locations. In fact, the Leap Motion stylus based calibrations yield far greater extrinsic estimates compared to both the finger alignments employed in this study and also compared to extrinsic values found in Study 1 and nearly all previous studies evaluating SPAAM procedures for environment-centric alignments.

The significantly more accurate extrinsic values for stylus alignments may be related to the tracking ability of the Leap Motion and the easily discernible tip of the stylus itself. The accuracy of hand tracking, by the Leap Motion, is highly dependent on the orientation, position, and occlusion level of the fingers, palm, and hand features. This inherent systematic variability naturally leads to less consistent measures in all three of our finger based alignment methods. Conversely, the high precision in stylus tracking inherently promotes more reliable results. Additionally, the presence of actual misalignment error between the target points of the on-screen reticles and the user's finger tip further increases the potential for inaccuracies and high variability in calibration results. Even so, results do show that the improved positioning context afforded by the box and finger reticles do positively effect calibration results for both monocular and stereo procedures, though not to a high enough degree to be comparable to the stylus results.

The heightened performance of the stylus calibration, compared to the results from prior investigative studies, is most likely a product of the environment-centric methodology employed in those systems. The SPAAM procedures employed, and evaluated, almost entirely utilize static locations or markers within the environment as alignment points. The systemic errors due to measurement inaccuracies of these alignment points is not present in the user-centric approach of this work. Also, all calibration alignments in these previous investigations were performed by standing users, which, as shown by Axholt et al. [8], increases the occurrence and magnitude of misalignment error due to postural sway. The Leap Motion calibration allows alignments to be performed while seated, reducing the tendency of sway and body motion during the procedure.

This investigation clearly shows that Leap Motion facilitated SPAAM calibration, in both monocular and stereo modalities, is able to yield result qualities well within acceptable levels, as compared to those presented in prior SPAAM evaluation studies. Additionally, the findings indicate that hand based calibration accuracy can be improved by using more visibly contextual reticle designs to aid in finger placement during alignment. Nevertheless, the higher tracking accuracy and repeatability of stylus alignments makes this the recommended method. Though the ultimate goal for user-centric calibration will be the removal of dependency on physical alignment targets, the inclusion of a storable stylus, in forthcoming consumer OST devices, is a reasonable requirement to facilitate manual calibration techniques, such as those explored in this work. The results also provide a conclusive reference for researchers and system designers wishing to implement a similar user-centric calibration design.

4.3 Study 3: Implementing User-Centric Calibration for Environment-Agnostic OST AR Systems

The results of Study 2 validate the feasibility of using a user-centric calibration methodology for OST HMD systems through hand or stylus tracking technology, such as a Leap Motion controller. This effort continues to build on the same line of work by realizing a completed design and implementation for a fully working environment-agnostic OST AR setup. Motivation for constructing such a system arises from the lack of existing material, content, and documentation on straightforward OST HMD methods and procedures using consumer level sensors that are also well suited for ubiquitous deployment across a wide range of environment types. The final goal of this endeavor is, therefore, to not

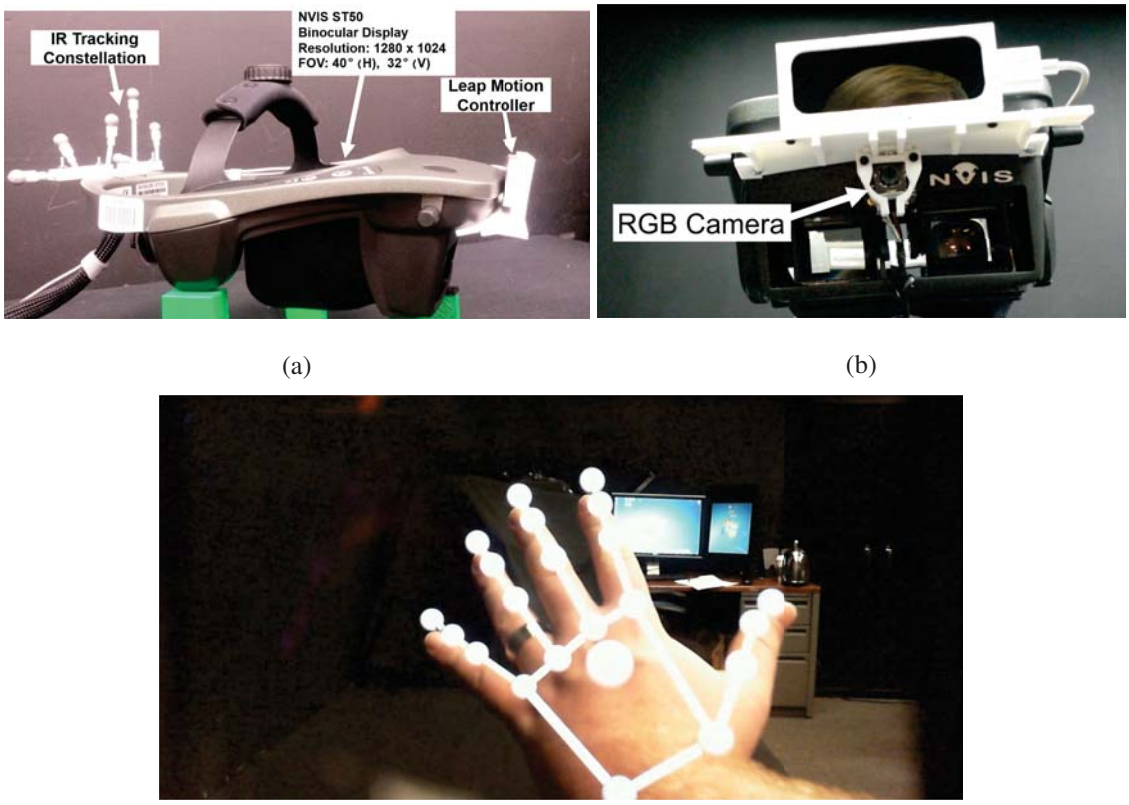
only illustrate a working environment-agnostic OST HMD system, but to also produce a usable framework that is easily replicated to inspire and encourage further developmental efforts towards standardized approaches for addressing the calibration needs of OST systems within the AR community at large.

4.3.1 User-Centric OST HMD Setup

A product of the work in Study 2 was the creation of a combined apparatus comprised of both a binocular OST HMD and a Leap Motion controller. This work continues to build on that existing hardware setup. The same NVIS ST50 OST HMD is again fitted with a front-facing Leap Motion controller attached using a custom 3D printed mount. This minimal hardware system is all that is necessary for the utilization of the user-centric calibration scheme presented in Study 2. Calibrating this setup using the described methodology allows for the deployment of AR applications capable of registering content to any object trackable by the Leap Motion. This includes hands, fingers, arms, and stylus tools.

Unfortunately, this setup alone does not provide the necessary tracking capabilities to extend the registration of content out into the world. For instance, there is no IMU or secondary tracking mechanism natively built into the HMD itself. Fortunately, the primary strength of adopting a user-centric system approach is the ability to easily expand the tracking capabilities of the setup without the need for altering the underlying calibration methodology. Figure 4.14 provides illustrations of how the hardware setup from Study 2 can be combined with two completely different types of secondary tracking technology: an RGB camera for inside-out fiducial marker tracking, Figure 4.14 (b), and a retro-reflective

constellation for outside-in optical tracking by IR cameras, Figure 4.14 (a). Figure 4.14 (c) provides a view through the system of a simple rendered skeletal overlay registered to a user's hand.



(a) (b)

(c)

Figure 4.14 Study 3 system hardware

(a) The NVIS ST50 with attached Leap Motion controller and IR retro-reflective markers
(b) The NVIS ST50 with attached Leap Motion controller and Microsoft Lifecam HD-6000 RGB camera
(c) Calibration results used to display a skeletal overlay onto the user's hand

4.3.2 Ubiquitous Deployment Through Leap Motion Coordinate Calibration

While the secondary tracking mechanisms shown in Figure 4.14 (a) and (b) may be used to generate 6 DOF position and orientation information about the user within the new

environment, this tracking information cannot be used directly by the AR application itself until the relationship between the new tracking frame of reference and that of the Leap Motion is established. Since the application running on the system has been calibrated to the Leap Motion coordinate frame, the secondary tracking input must be transformed into this same frame of reference before being utilized for positioning virtual content.

The Leap Motion device itself is internally a set of stereo IR cameras. Using existing computer vision techniques [145, 46] it may be possible to utilize the visual information from both cameras to create depth maps or point clouds of the surrounding environment. This strategy, though, is not completely viable due to the limited visible range of the Leap Motion cameras themselves, and also because of the level of knowledge and added implementation costs required to implement the necessary image processing algorithms. In order to correspond with the stated goals of this work, a novel, extremely low cost, easily accessible calibration method is devised to determine the transformation between the Leap Motion coordinate frame and nearly any secondary tracking system.

Consider again the tracking modalities shown in Figure 4.14 (a) and (b). Calibrating an RGB camera coordinate system would require a jig using a visible fiducial marker. Likewise, an optical IR camera system would require a passive retro-reflective marker ball. Incorporating these requirements with a stylus, similar to that used for the OST calibration scheme, will make it possible to leverage point information from each tracking source to accomplish the transformation task.

The tool tracking ability of the Leap Motion makes it possible to natively acquire a large amount of information about a stylus object. The tip of the stylus, for example, is

used in Study 2 for providing the 3D world correspondence point data required by the SPAAM algorithm. In addition to the tip 3D position, the Leap Motion is also able to provide the pointing direction of the stylus as a 3D vector with an X, Y, and Z directional component. Provided a tip and pointing direction, it is therefore possible to determine the 3D position of another object located along the length of the stylus. Figure 4.15 (a) and (b) show two physical calibration jigs created around this principle to provide location references to identical points within the Leap Motion and the secondary coordinate frames of the tracking systems from Figure 4.14.

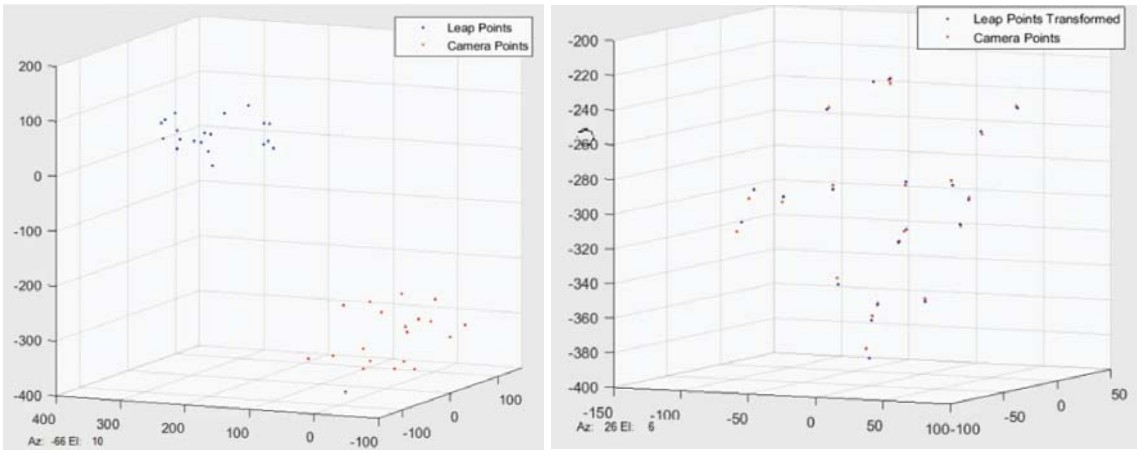
(a) (b)

Figure 4.15 Leap Motion calibration jigs

(a) Coordinate frame calibration jig comprised of an aluminum rod and visible fiducial tracking marker
(b) Coordinate frame calibration jig comprised of an aluminum rod and retro-reflective IR tracking marker

The calibration jig shown in Figure 4.15 (a) is constructed such that the center of the fiducial marker lies along the center line of the stylus. The length of the rod can be readily measured through a number of highly accurate means, in this case it was machined to a length of approximately 10cm, with a tolerance of $\pm 1mm$. The fiducial marker itself is affixed to a custom 3D printed holder designed so that the center of the mounted marker

is approximately 3.5cm from the end of the stylus. Using the free end of the stylus as the tracked tip, reversing the pointing direction provided by the Leap Motion, and the known tip to marker center distance of the jig itself, it is possible to acquire two sets of point data relating the 3D position of the fiducial marker's center. The first set describes the location of the marker center in relation to the RGB camera, and the second the location of the marker center in relation to the Leap Motion coordinate frame. Figure 4.16(a) provides a 3D plot of the two point sets.



(a) (b)

Figure 4.16 Point cloud data sets representing the 3D location of the fiducial marker center

Locations are with respect to the coordinate frame of the Leap Motion (blue) and RGB tracking camera (red)
(a) Point sets before transformation
(b) Point sets after applying the transformation result from the Absolute Orientation calculation

The goal of this procedure is to determine the transformation between the two coordinate frames. This process is described by an absolute orientation operation, which is able to be solved using a number of possible techniques [2, 61, 149]. The methodology

employed in this system is that proposed by Shinji Umeyama [149], which uses a least squares methodology to determine the transformation that best fits the values from one point set to the locations in the second. Figure 4.16 (b) shows the same point sets from Figure 4.16 (a) transformed by the absolute orientation solution. The accuracy of the solution is, of course, dependent on a number of factors including the precision to which the physical jig is constructed, the reliability of the tracking data used to create the point sets, and the number of points used for the operation.

This entire procedure is able to be performed off-line, and is only required to be performed once granted that the relative positions of the Leap Motion and the secondary tracking reference do not change during use. Once the required coordinate frame transformation is known, the tracking data provided by the secondary tracking system can be transformed into the Leap Motion frame of reference allowing the AR application to appropriately position virtual content according to the new environment, and without the need to adjust the calibration of the OST display itself. This complete procedure is available as a Technical Video, with the accompanying abstract provided in [94].

4.3.3 Working Demonstration System

In order to showcase the versatility of a user-centric system design, an immersive stereoscopic AR experience is constructed allowing users to both calibrate the HMD and perform natural interaction with virtual objects registered to the environment using the tracking data provided by a Leap Motion and a secondary tracking system. The complete hardware setup for the application is shown in Figure 4.14 (a), combining the HMD and

Leap Motion apparatus from Study 2 with tracking capabilities from an externally mounted IR optical tracking unit.

4.3.3.1 Software and Hardware

6 DOF head tracking data is provided by an ART Trackpack camera pair with version 2.10.0 of the accompanying DTrack2 software. As in Study 2, the Leap Motion tracking information is acquired using version 2.3.1.31549 of the Leap Motion SDK. The application used to control the rendering and interaction of virtual content is written in C++ utilizing an OpenGL based pipeline. The position of the Leap Motion relative to the IR constellation, mounted on the rear of the HMD, is determined prior to run-time using the previously described absolute orientation methodology adapted to use the physical jig shown in Figure 4.15 (b).

4.3.3.2 User Interaction

The immersive application affords users two forms of direct interaction with our system: calibration of the HMD, and participation in a target striking game. Calibration of the OST display is performed on-line by the user using an identical procedure to the stereo stylus condition described in Study 2. The calibration and registration quality may be examined by the user through examination of a simple skeletal overlay, shown in Figure 4.14 (c), or by participation within a simple target striking game.

Figure 4.17 illustrates the user's presence within the complete demo application. During the target game, participants are able to use the stereoscopic cues, provided by the binocular display, to grab a small virtual ball, which they may then toss in an effort to

strike one of several virtual targets arrayed before them. The benefit of utilizing a Leap Motion controller for use in the system is further highlighted by the ability to produce simple occlusion of the virtual objects by the user's hands using the tracking data from the Leap Motion as a reference for depth buffer checking. The complete system has been showcased at the 2016 IEEE Virtual Reality Conference in Greenville, South Carolina, [95].

Figure 4.17 Demonstration application

(**Left**) A participant in full swing, engaged in the target game portion of the application (**Top Right**) View through the HMD of the menu selection within the demonstration (**Bottom Right**) Hand interaction used to toss the ball at targets within the game

4.3.4 Discussion and Conclusion

The final goal of this study is to not only illustrate a working environment-agnostic OST HMD system, but to also produce a usable framework that is easily replicated to inspire and encourage further developmental efforts towards standardized approaches for addressing the calibration needs of OST systems within the AR community at large. In addition to a live use-case of the user-centric calibration approach evaluated in Study 2, this work is also the first to present a novel calibration scheme for determining the transformation between a Leap Motion's coordinate frame and that of a secondary tracking system. This coordinate frame calibration method allows the Leap Motion and HMD system to be easily and readily

incorporated into nearly any application environment without the need for retailoring or altering the HMD calibration procedure itself. Additionally, the procedures and related content for the system outlined in this work have been made available to the AR community through both a Technical Video and live demonstration accomplishing the stated goals of disseminating a usable reference onto which further improvements can be made and a plethora of innovative and novel experiences created.

4.4 Study 4: Improved Stereo Calibration Through Nonious Visualizations

While the results of Study 2 and 3 show that user-centric calibration of OST HMD's is a viable, and more versatile, approach compared to typical environment-centric approaches, there still remains a large number of systems that will be incapable of or inhibited by adopting an environment-agnostic process. These systems, however, will still benefit from the utilization of a stereo calibration scheme, which has been shown in Study 2 to provide significant improvements in consistency, accuracy, and robustness compared to monocular variants. While implementing a stereo SPAAM procedure is relatively straight forward, there are a number of factors that may inhibit the usability and impact the accuracy of the implementation.

As discussed in Study 2, stereo SPAAM approaches leverage binocular HMDs to create a perception of depth in the virtual content. This is, of course, accomplished by rendering different images to the left and right eyes to induce stereopsis in the fused image. Unfortunately, most all current generation OST HMDs are only able to render virtual content at a fixed focal distance. This means that even though the stereo cues from the display will

cause the user's eyes to converge as though they were viewing objects at varying distances, their accommodative demand, or focal ability, will remain constant. This convergence–accommodative rivalry is especially noticeable when virtual and real world objects are viewed simultaneously, as is the case during the alignment process of the stereo SPAAM calibration procedure. This work is motivated by the need to create effective strategies for addressing the possibility of accommodative–convergence mismatch in stereo SPAAM implementations that are also applicable across a wide range of OST HMD systems.

During a SPAAM alignment, the user will attempt to align the on-screen reticle with the 3D point in the world. When the accommodative and convergence cues between the reticle and world point clash, double vision, focus instability, and eye strain can result, making the process all the more tedious and fatiguing. These accommodative–convergence mismatch effects can be ameliorated to an extent by ensuring that the on-screen reticle is also rendered at a depth able to be matched by the world point. For example, in Study 2, the reticles for the stereo calibration conditions were rendered to ensure that their perceived depth was always within arms reach, allowing the hand-held stylus to simply be placed at the appropriate distance for the correspondence.

Unfortunately, the ability to control the perceived depth of on-screen reticles may not always be possible, or it may also be the case that the distance to the world point to be used is not known before hand. Creating pre-defined reticle placements for these conditions will be difficult, and it may often be an easier solution to simply adopt a monocular, one eye at a time, calibration approach. This alternative, of course, sacrifices the accuracy gains shown for stereo calibration for implementation ease.

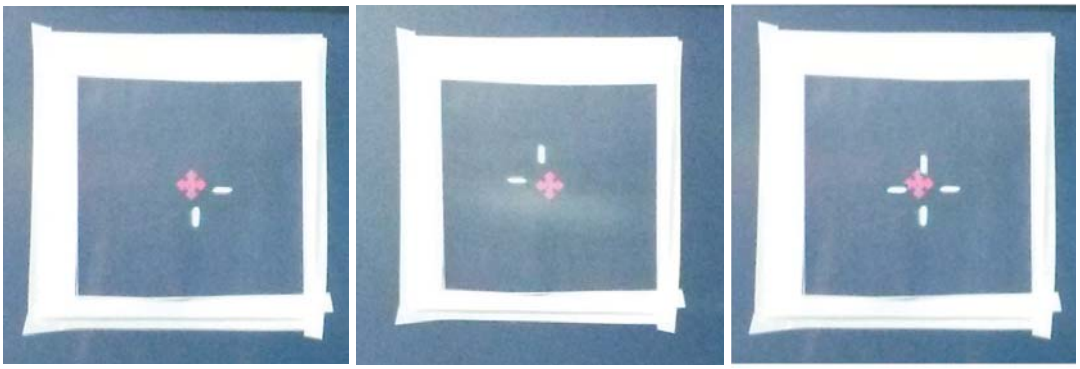
This work provides an alternative method for reticle design, based on the concept of nonius lines, that is well suited for use in environment-centric stereo SPAAM implementations. Additionally, the results of a preliminary follow-up investigation to Study 2 examining the consistency of a standard monocular and a stereo SPAAM implementation using the new reticle design is presented. Though a description of the experiment procedure and results follow, the published format of the work is available in [97].

4.4.1 Nonius Reticles

The use of nonius lines for measuring the stereo vergence angle of humans is well documented and an often employed technique for diagnosing stereo blindness and other optical abnormalities [26, 88, 120, 127]. Implementation of a nonius line visual is also rather straightforward. Similar to standard stereo images which fuse into a single object at an apparent depth, nonius lines are simple pairs of vertical line segments that will align, appear to be collinear, when viewed with a certain eye vergence angle. By shifting the position of one line left or right, the required vergence angle to fuse the two segments into a contiguous line will also change. This work applies this same methodology to create a nonius reticle style for stereo SPAAM calibrations.

Standard stereo SPAAM implementations commonly employ solid reticle designs, such as that shown in Figure 4.18 (c), or those used in Study 2. As noted previously, the ability to fuse these reticles through stereopsis is often inhibited by the accommodative–convergence rivalry that arises from viewing on-screen and world objects simultaneously. The improved reticle design investigated in this study splits the on-screen object into two distinct halves,

shown in Figure 4.18 (a) and (b). This new nonius reticle allows users to focus solely on a point in the environment and adjust the on-screen locations of the reticle halves until they are perceived as being properly aligned. This approach eases the burden on the system developer by removing the need to create predetermined screen positions for reticles to force alignments at certain depths. Similarly, it reduces the visual strain on the user by significantly removing the possibility for accommodation–convergence mismatch during the alignment process.



(a) (b) (c)

Figure 4.18 Views through the HMD of the alignment marker and crosshair

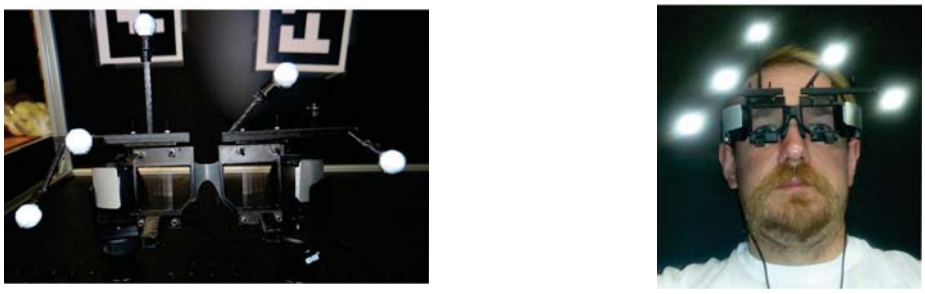
(a) Left eye nonius cross-hair half
(b) Right eye nonius crosshair half
(c) Full cross-hair

4.4.2 Preliminary Experiment

Based on the results from Study 2, which show a marked improvement in accuracy for stereo compared to monocular calibration, the efficacy of the nonius reticle for facilitating stereo calibration is examined through a small preliminary study.

4.4.2.1 Hardware

A Lumus DK-32 HMD is used as the primary display for this study. This is a binocular OST HMD with a resolution of 1280×720 per eye and 40^o diagonal field of view. A passive IR marker constellation is rigidly attached to the upper edge of the HMD frame, Figure 4.19. The constellation's position and orientation are tracked using a pair of ART Trackpack cameras, with a resolution of .7 MPix and a 90Hz update rate.



(a) (b)

Figure 4.19 Components of the Study 4 calibration system

(a) The Lumus DK-32 head mounted display with tracking constellation
(b) User wearing the Lumus DK-32 with illuminated constellation

4.4.2.2 Calibration Procedures

Similar to Study 2, this investigation compares the performance of a monocular and stereo SPAAM calibration variant. Both conditions are conducted in an identical manner with only the display style of the on-screen cross-hair differing between the two. Unlike study 2, only 20 screen–world alignments are performed for each completed calibration trial. During alignment, the user is instructed to line up the center of the on-screen cross-hair with the center of a physical marker. The user is instructed to take steps forward or backward between alignments to vary the distance of each measurement.

During the monocular SPAAM condition all 20 alignments are made with only a single eye, right or left, and then the calibration repeated for the remaining eye. The unused eye is covered to avoid binocular rivalry during alignments. Figure 4.18 (c) provides a view of the on-screen cross-hair shown to the user during alignments.

The stereo SPAAM condition proceeds nearly identically to the monocular case except that the nonius reticle design is utilized instead of a solid cross-hair. As described previously, half of the cross-hair is displayed to each eye simultaneously, Figures 4.18 (a) and (b). The user's optical system then fuses the two halves into a single image. During alignment, the user is instructed to focus on the center of the physical marker, then using a controller, independently adjust the on-screen location of each cross-hair half until the vertical and horizontal portions align to form a fused cross-hair image.

4.4.3 Results

As in Study 2, a single expert user provided all of the calibration data. A total of 5 stereo and 10 monocular (5 for each eye) SPAAM calibrations were performed total. The evaluation metric used is the same binocular disparity measures employed in Study 2, taken as the difference between the left and right eye location estimates along each major direction: horizontal (IPD), vertical, and in depth. Figures 4.20 (a), (b), and (c) shows the value of the median eye position differences along each direction after each alignment of the calibration procedure. Negative values indicate the right eye position estimate is greater in the indicated direction. For each disparity, stereo SPAAM, plotted in blue, not only achieves a steady state value sooner than the monocular SPAAM condition, but also

exhibits significantly less deviation between the first and last estimates. The change in IPD during calibration for stereo SPAAM is only .89cm compared to 2.34cm for monocular SPAAM. Stereo SPAAM also varies by only 0.09cm and 1.34cm compared to 0.99cm and 6.47cm for monocular SPAAM in the vertical and depth directions respectively.

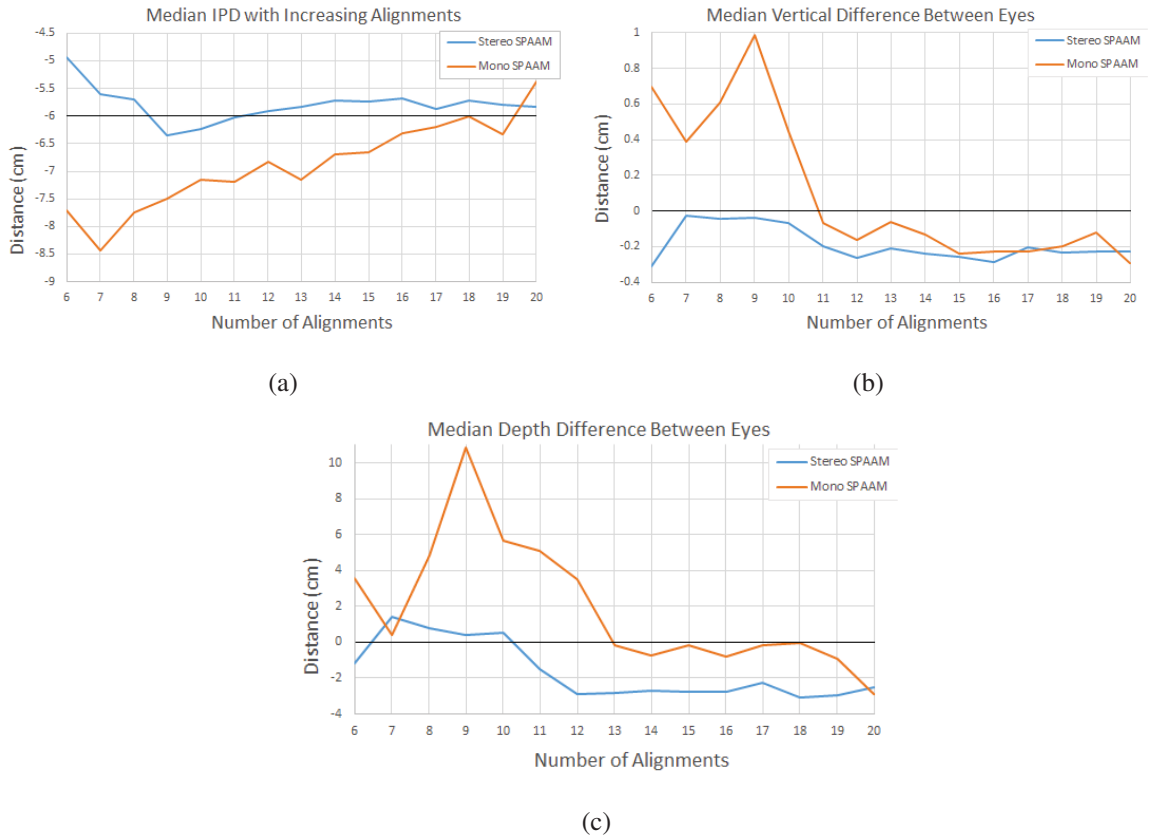


Figure 4.20 Plots of the median 3D binocular disparity measures

Calibration conditions are denoted in blue (stereo) and orange (monocular)
(a) Horizontal eye position difference (IPD)
(b) Vertical eye position difference
(c) Depth eye position difference.

Accuracy of the calibration can be evaluated using the IPD as an indicator. Mean IPD estimates for calibration are approximately 5.8cm and 6.8cm for stereo and monocular

SPAAM respectively. The stereo SPAAM estimate is closer to the user's real IPD value of 6cm and also maintains this value through nearly the entire calibration process, where as monocular SPAAM estimates slowly improve and only approach the correct value near the calibration end.

4.4.4 Conclusion

The preliminary results provided in this study indicates that the stereo SPAAM method, using the nonius reticle design, is able to produce more consistent results compared to the standard monocular variant. These results align well with those found in Study 2 for the user-centric stereo SPAAM condition, as well as the improved robustness for stereo vs monocular calibration. This study also provides a look at the performance of the SPAAM calibration over the last 14 alignments, since a minimum of 6 is actually needed to acquire a minimal solution. These findings confirm the viability of the nonius reticle approach and encourage application of the method in future expanded evaluation studies examining the performance of multiple users, or the convergence of user-centric and environment-centric implementations with increasing alignment counts.

4.5 Study 5: Frustum Visualization as an Evaluation Alternative

Even though the evaluation metrics employed in Studies 1-4 are able to provide adequate measures for the quality of a calibration result, there still remains the limitation that only the user himself is able to actually view the result through the display. While subjective evaluation tasks, such as those used in Study 1, provide additional feedback with regards to registration error, the utility and trustworthiness of this information is greatly

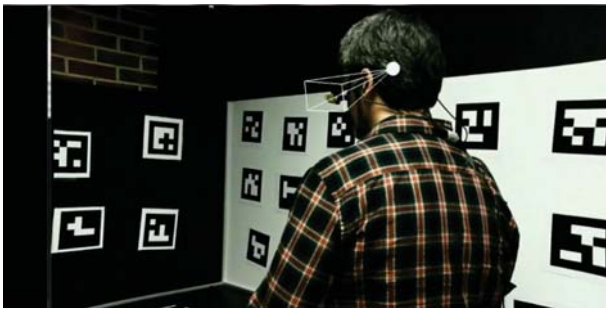
dependent on the subject's ability to perform the task effectively. This work describes the result of an exploratory investigation into an alternative method for visualizing the user's view through the display, that would allow a third party observer to see from the user's perspective within the system. An explanation of the technique is available as a Technical Video at the 2015 IEEE Virtual Reality Conference, with the published abstract provided in [96].

4.5.1 Frustum Generation

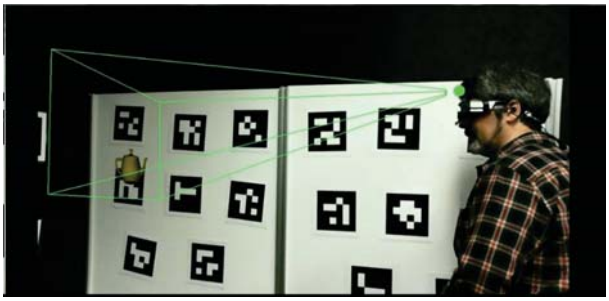
The results of OST HMD calibration explicitly describe a model for the user's perspective through the device. As described in section 3.2.1, this result is actually comprised of two components, a set of extrinsic parameters which describe the location of the user's eye relative to the HMD coordinate frame, and a set of intrinsic parameters which describes the shape of the viewing frustum. By separating these two components from the calibration result, it is possible to generate a visualization of the viewing frustum as it would appear to a third party observing the user.

Figure 4.21 provides a view of an example system which implements this strategy. A user, wearing a Lumus DK-32 HMD performs a standard monocular SPAAM calibration, where the location of the world alignment point is obtained through standard visual fiducial marking tracking by an RGB camera mounted to the HMD frame. During the calibration process, the user performs more and more alignments, which update the resulting projection matrix. This projection matrix is decomposed into the extrinsic and intrinsic parameters after each update from a new alignment pair. A model of the viewing frustum

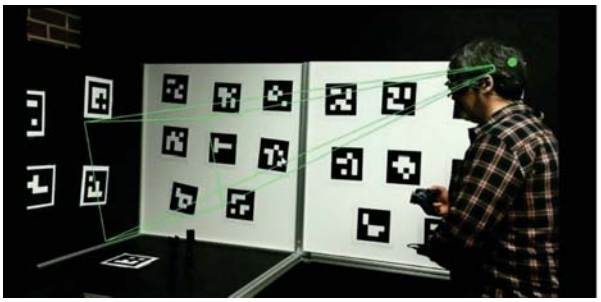
can then be created based on the intrinsic values. A third party observer, tracked within the same world coordinate frame as the user, may then view the modeled frustum collocated to the position and orientation described by the extrinsic parameters of the calibration result.



(a)



(b)



(c)

Figure 4.21 Frustum visualization of a SPAAM calibration

(a) (b) Visualizations of a user's SPAAM results with on-screen geometry overlayed onto the far plane of the viewing frustum
(c) Images of a user performing screen-world alignments during SPAAM calibration

The visualized frustum is able to provide the observer with immediate feedback with regard to the accuracy of the extrinsic values as well as the shape of the viewing frustum. Further enhancements to the visualization can be made by rendering the on-screen content to an additional framebuffer, which is then mapped to the frustum. This augmented view would allow the observer to not only see the content being viewed by the user, but also its relative location within the world from the user's perspective, Figure 4.21 (b). Though the visualizations provided in Figure 4.21 are for a single eye's view, the same methodology could easily be extended to produce simultaneous visualizations for the left and right viewing frustums of the user.

4.5.2 Application of the Technique

A similar visualization strategy has been employed for marketing next generation OST HMDs, by showing a rendering of what a user is seeing from the vantage point of a third party observer, Figure 4.22. This methodology, however, does not provide any indication that the visuals are actually collocated to where the user perceives them to be. This use case is most similar to a telepresence, or remote collaboration environment, in which both parties are able to observe virtual content, but the perceived location of the content by one user does not affect the efficacy of the view from the second user. Instead, the frustum visualization approach described within this work explicitly defines the intersection of the user's view with the environment, allowing for a direct examination of the apparent registration error by the third-party observer.

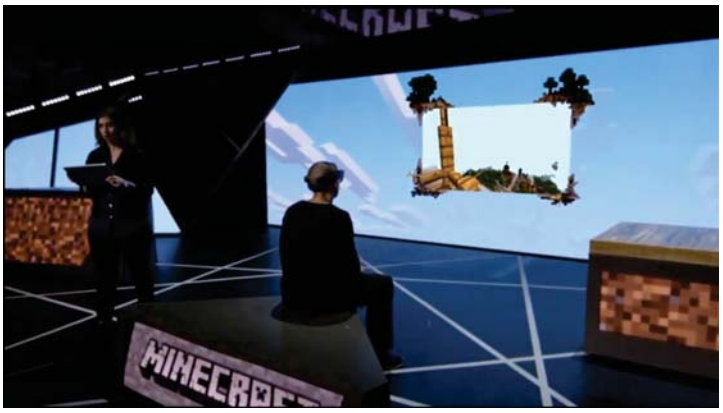


Figure 4.22 Visualization used during a demonstration of a Microsoft Hololens

The virtual screen is rendered onto an external video feed to provide the audience an indication of what the demonstrator is seeing through the device

This frustum technique can also be expanded further when a 3D model of the environment is available. Provided that layout and geometry information of the user's space is known, it will be possible to recreate the environment in a virtual space and, provided that the user's head and movement within the environment is tracked, allow the third party observer to see directly from the user's eyes during run-time. The utility of this approach actually extends beyond evaluation of HMD calibration quality and would also be applicable for telepresence and remote collaboration applications as previously noted.

4.6 Study 6: Direct Comparison of User-Centric and Environment-Centric Calibration Accuracy

This final work consists of a follow-up study intended to address and examine several questions produced by the results of Study 2 and 3. The outcomes of Study 2 show that the user-centric calibrations performed using a stylus results in extrinsic eye location estimates that are significantly more consistent and accurate, relative to plausible eye locations, compared to nearly all previous studies evaluating environment-centric SPAAM implementa-

tions. A plausible explanation for the increase in performance of the stylus method is the high degree of tracking accuracy provided by the Leap Motion compared to the modeled accuracy of the alignment points used in previous investigations. Likewise, the stylus alignments were performed while the user was seated, greatly decreasing the production of postural motion, though this mechanic should have logically also improved the results of the finger alignment procedures as well. Another explanation is that the alignment points used for Study 2 were taken at near-field, arms length, distances, which contrasts with the range of alignment points implemented in prior work. This study investigates these issues further using a modified version of the Study 2 experiment, expanding the conditions to include not only user-centric and environment-centric modalities, but also a control condition designed to remove degrading alignment effects from user postural sway.

4.6.1 Experimental Design

It is possible to mimic the user-centric calibration methodology from Study 2 using an outside-in IR optical tracking system. In this revised system, the Leap Motion is replaced by an affixed retro-reflective constellation, such as the one used in Study 3. This modification would also mean, though, that the stylus tracking would need to be simulated as well. This would be possible by replacing the stylus with a hand-held retro-reflective marker also tracked by the same system. Using an identical stereo methodology, the calibration results for this system should closely match those seen for the stylus condition of Study 2. Deviations between the results of this second study and those from Study 2 would indicate

an influence from the tracking system itself, most certainly due to accuracy differences in measuring the alignment point location.

Further comparison is also facilitated by extending the stereo calibration to use alignment points at the user-centric, arms length, distances from Study 2, as well as medium-field distances, such as those employed for the SPAAM implementation of Study 1. The same IR optical tracking system can be utilized for this extension as well, in order to maintain consistency in systemic tracking errors. The same hand-held retro-reflective marker is simply affixed to a tripod approximately .5m–2m from the user. While it is expected that the alignment range itself will have the largest influence on alignment accuracy, movement of the user during the environment-centric condition, as discussed in Magnus Axholt's prior work, may also heavily contribute to degraded performance in this condition.

In order to effectively ameliorate postural sway influences, two sets of stereo calibrations are performed for both user-centric and environment-centric alignment ranges. The first set of calibration data is obtained from a seated user, with the second data set produced by the same user while standing. Deviations between the two results will indicate a greater contribution due to postural sway during the alignment process. It is important to note that deviation between the medium and near field calibration results may arise from inherent human performance limitations for performing alignments to world points at any significant range. A control condition, therefore, is also implemented to remove any influence from human motion and motor control from the alignment procedure. In order to accomplish this, the HMD itself is rigidly mounted to an adjustable tri-pod, which is then manually maneuvered to perform the alignment procedure. The combined sitting/standing condi-

tions for the user at each of the two alignment ranges, user-centric/environment-centric, coupled with the control condition also performed at the two alignment ranges yields a total of six final conditions investigated and compared by this study. This work, therefore, is not only the first formal evaluation explicitly aimed to compare the results of user-centric and environment-centric SPAAM calibration modalities, but is also the first study to implement a novel control condition to provide base-line SPAAM results devoid of error from subjective motor control limitations.

4.6.1.1 Hardware System

The same NVIS ST50 binocular OST HMD, used in all prior studies, is also used as the primary display for this investigation. The full binocular capabilities of the display are also utilized during all conditions in which a user is present. A custom 3D printed mount is used to attach a retro-reflective marker array to the front of the display, Figure 4.23. This constellation is used, in conjunction with an ART Trackpack camera system, to provide 6 DOF position and orientation data for the HMD within the experimental environment. Unlike Study 2, Leap Motion tool tracking is not available for this investigation. Instead, the physical target point for all conditions of this study is taken to be the center of a 6mm diameter retro-reflective sphere attached to the end of a cylindrical rod. The 3D position of the sphere is actively measured during calibration by the ART tracking system and used in combination with the HMD pose to determine the head relative coordinate of the target during the alignment procedure.

Figure 4.23 ST50 HMD with retro-reflective constellation

Additional hardware is also utilized to create the user-absent, control, condition. This assembly is designed to mimic the view of a user through the display but also provide a mechanism for performing the necessary calibration alignments without the presence of postural motion. In order to accomplish this, the HMD itself is rigidly mounted to a camera-tri-pod system. A Microsoft Lifecam HD-6000 webcam, with a resolution of 1280×720 at 30fps, is mounted within the display using an optical railing system. The camera is able to be adjusted in 4 DOF, vertical, horizontal, lateral, and yaw, to provide a view through the HMD screen at an approximate location that a user's eye would naturally occur. The entire HMD and camera system is also mounted to a tri-pod and geared adjustment head to allow for movement and rotational alignment of the entire HMD assembly for calibration. Figure 4.24 provides views of the camera and HMD mounting system for the user-absent condition. The view from the camera is captured through a USB 2.0 connection to a secondary laptop running the Microsoft Lifecam software.

(a) (b) (c) (d)

Figure 4.24 Camera system for user-absent condition

(a) Top view of HMD and camera system
(b) View of camera behind the HMD screen
(c) Side view of the camera optical rail mounting
(d) view of the tri-pod and gear head assembly

4.6.1.2 SPAAM Procedure

A standard manual SPAAM calibration procedure is employed for this study, as described in [148]. Normalization of the 2D screen and 3D world points is also incorportaed into the procedure as recommended by [52]. During the calibration, the participant is provided an on-screen reticle and is tasked with aligning the center of the reticle with the center of the physical target point, previously described. One of two types of on-screen reticles is employed depending on calibration condition. The user-centric and user-absent calibration conditions employ the same Cross reticle utilized in Study 2. This reticle is a simple cross-hair comprised of a horizontal and vertical line with the alignment point located at the center of the reticle. The on-screen dimensions of the cross are 64×64 pixels

with line thickness of 3 pixels. The on-screen reticle for the environment-centric calibration condition employs the nonius reticle design discussed in Study 4. The Cross reticle is separated into two halves, with one half shown to each eye. The participant is able to adjust the on-screen location of the right half of the cross-hair until the two halves appear to visually align into the complete cross at the physical target point. Figure 4.25 shows a view through the HMD of a visual alignment between cross-hair and target point. A total of 50 alignments are performed to complete a singular calibration set within each condition. The distance separation between the participant and the target point varies according to the calibration modality: user-centric or environment-centric.

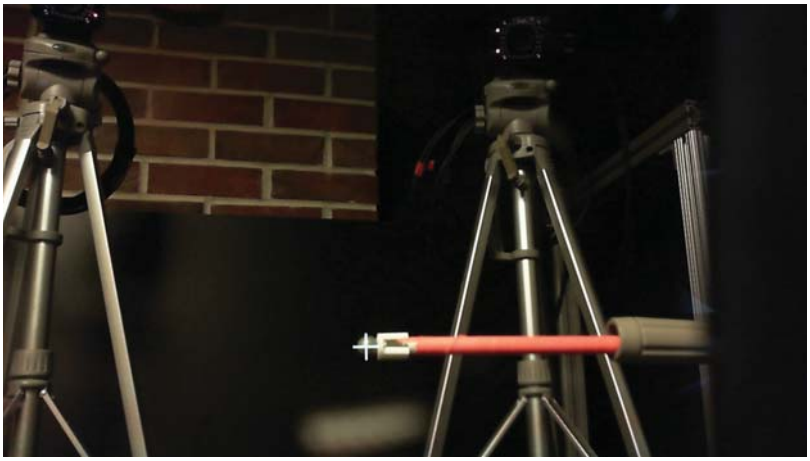


Figure 4.25 View through the HMD of reticle to target alignment

4.6.1.3 User-Centric Alignment Distances

The user-centric calibration condition employs a nearly identical methodology to the stylus calibration procedure utilized in Study 2. The user is presented the on-screen cross-hair, positioned in each eye to induce stereopsis and the perception of the cross-hair in depth. The binocular placement of the cross-hair on-screen is controlled such that the

perceived depth of the reticle extends in-front of the user between .15m and .3m, or approximately arm's length, with the distance a each alignment described by a Magic Square distribution, as recommended by [3]. The participant performs the calibration procedure previously described in one of two stances: seated within a chair with back support and arm rests, or standing with no additional body support provided. Figure 4.26 provides illustrations of an example participant performing a user-centric calibration in both conditions.

(a) (b)

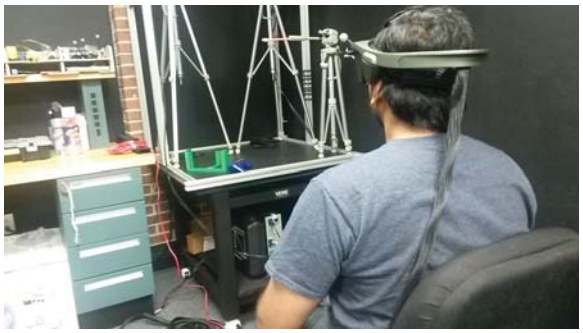
Figure 4.26 User-centric calibration condition

(a) Participant performing the calibration while seated
(b) Participant performing the calibration while standing

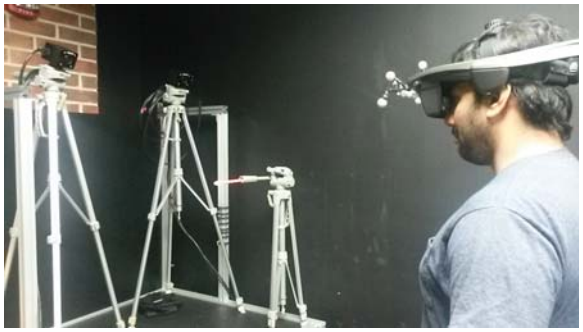
4.6.1.4 Environment-Centric Alignment Distances

The environment-centric calibration utilizes a similar configuration to that from Study 1 and most all prior studies investigating SPAAM calibration. The user is presented the on-screen cross-hair, using the nonius style previously discussed. The distance between the participant and the physical marker is varied between .5m–2m by the user taking steps forward or backward, or by adjusting the location of the chair toward or away, from the target point. The amount of distance varied between alignments is derived from a Magic Square distribution, as recommended by [3], with the distances marked along the ground

on a measured tape. The target point itself is affixed to a tripod and adjusted to the approximate height of the user and, as in the user-centric condition, the participant performs the calibration either standing or sitting. Figure 4.27 provides illustrations of an example participant performing an environment-centric calibration in both conditions.



(a)



(b)

Figure 4.27 Environment-centric calibration condition

(a) Participant performing the calibration while seated
(b) Participant performing the calibration while standing

4.6.1.5 Control User-Absent Condition

As stated previously, a control condition is also utilized in this study to compare calibration results from both user-centric and environment-centric alignments against identical calibration sets devoid of postural motion errors. This control, user-absent, condition uti-

lizes the camera and HMD tri-pod mounting system previously discussed. Identical sets of distances, .15m–.3m and .5–2m, are used for this condition in order to provide comparable calibration measures for both sets of alignment distances. During this condition, the monocular cross-hair, left eye image from the user-centric condition, is utilized, and the view from the webcam is referenced in order to adjust the orientation of the HMD to align the cross with the physical target point. While the process of adjusting the HMD orientation is performed manually, the precision of the alignment is still expected to far exceed that possible from a standard user-present calibration, since the postural and head motion from a user would cause a significant amount of pixel deviation from that attainable from the control apparatus.

4.6.2 Participant

All calibration data, with the exclusion of the control user-absent condition, is recorded from repeated trails by a single expert user, as in Study 2. Once again, the primary objective of this study is to compare the resulting accuracy of user-centric versus environment-centric calibration schemes. Restricting the calibration data to repeated measures from an expert user, knowledgeable with the procedure, removes the potential for errors resulting as an artifact from the subjective abilities of multiple participants. The expert subject completed 20 user-centric and 20 environment-centric calibrations in both a standing and sitting position, resulting in 20 x 2 x 2 = 80 calibrations total. The user-absent condition utilized 20 calibrations using the user-centric and environment-centric distance ranges for 20 x 2 =

40 additional calibrations. The user's maximum IPD is also measured to be approximately 62mm.

4.6.3 Study Results

As in Study 1 and 2, standard objective metrics for analyzing calibration accuracy are employed for this analysis. These measures include 3D eye location estimates and reprojection error. In addition to these values, an examination of the convergence, or trend over time, of these metrics is also included.

4.6.3.1 Eye Location Estimates

The estimated user eye location is taken from the extrinsic component of the projection matrix produced by the calibration. Figure 4.28 provides plots of the 3D eye locations resulting from calibrations performed in the user-centric, environment-centric, and user-absent alignment conditions. Two sets of plots are provided for each condition. The first shows the final result after the full 50 alignments. The second set shows the estimated locations after the first 25 alignments are performed. Through visual inspection, it is clearly evident that both the user-centric and user-absent conditions produce eye estimate values with far less variance compared to the environment-centric procedure, Figure 4.28 (b) and (e). Likewise, there is a prominent deviation between the user-centric and environment-centric variants of the user-absent, control condition, Figure 4.28(c) and (f). The user-centric alignment distances, in blue, are significantly more clustered and consistent compared to the eye estimates taken from the environment-centric, red, alignment results. In

contrast, the seated and standing participant data sets, plotted in red and blue respectively, of Figure 4.28 (a), (d), (b), and (e) do not exhibit much visual difference in values.

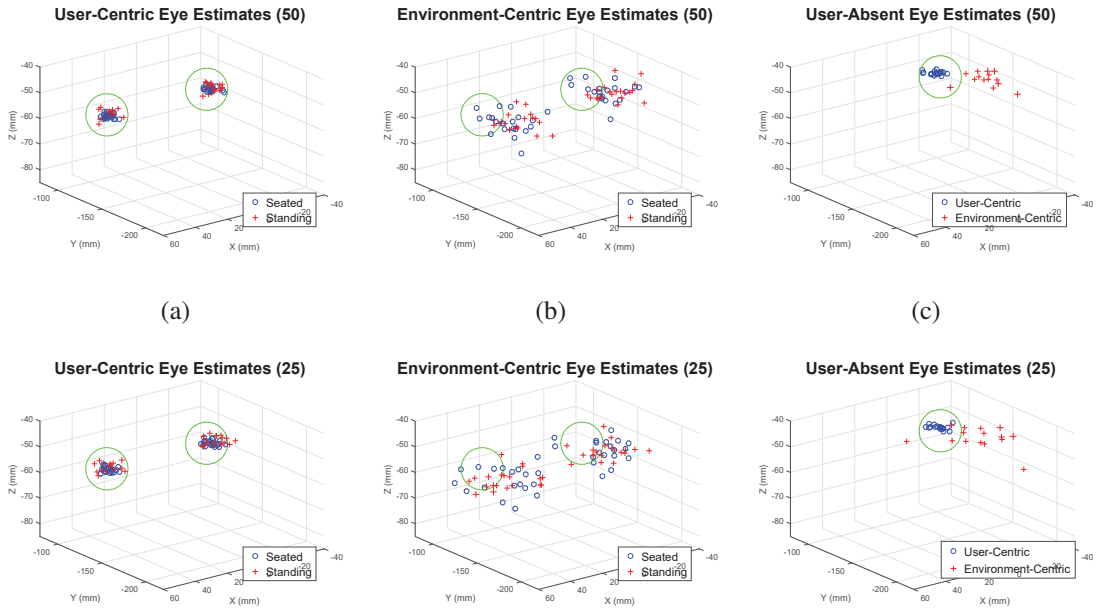


Figure 4.28 Estimated 3D user eye locations relative to the HMD marker constellation

(a) User-Centric, (b) Environment-Centric, and (c) User-Absent eye estimates after 50 Alignments. (d) User-Centric, (e) Environment-Centric, and (f) User-Absent eye estimates after 25 Alignments. In all plots, the center of the tracking constellation is at location (0, 0, 0). Seated user calibrations are displayed in blue, standing in red. Mounted user-centric calibrations are displayed in blue, with mounted environment-centric plotted in red.

Figure 4.29 shows a 2D cross-section of the eye estimate plots for each condition. The plots reiterate the visual clustering seen in the 3D graphs. Of particular note, however, is the trend of greater variability in the lateral, Y axis, particularly in the environment-centric condition, Figure 4.29 (b) and (e). This matches the recurring trends seen in Study 1 and prior work from Axholt et al. [10], which also employ environment-centric calibration modalities. Similarly, the user-absent environment-centric condition, red in Figure 4.29 (c)

and (f) also exhibits this identical propensity for increased lateral variance in eye estimates. The User-Centric states, for both the user Figure 4.29 (a) and (d) as well as user-absent Figure 4.29 (c) and (f) blue, show some lateral variance but not nearly to the same degree.

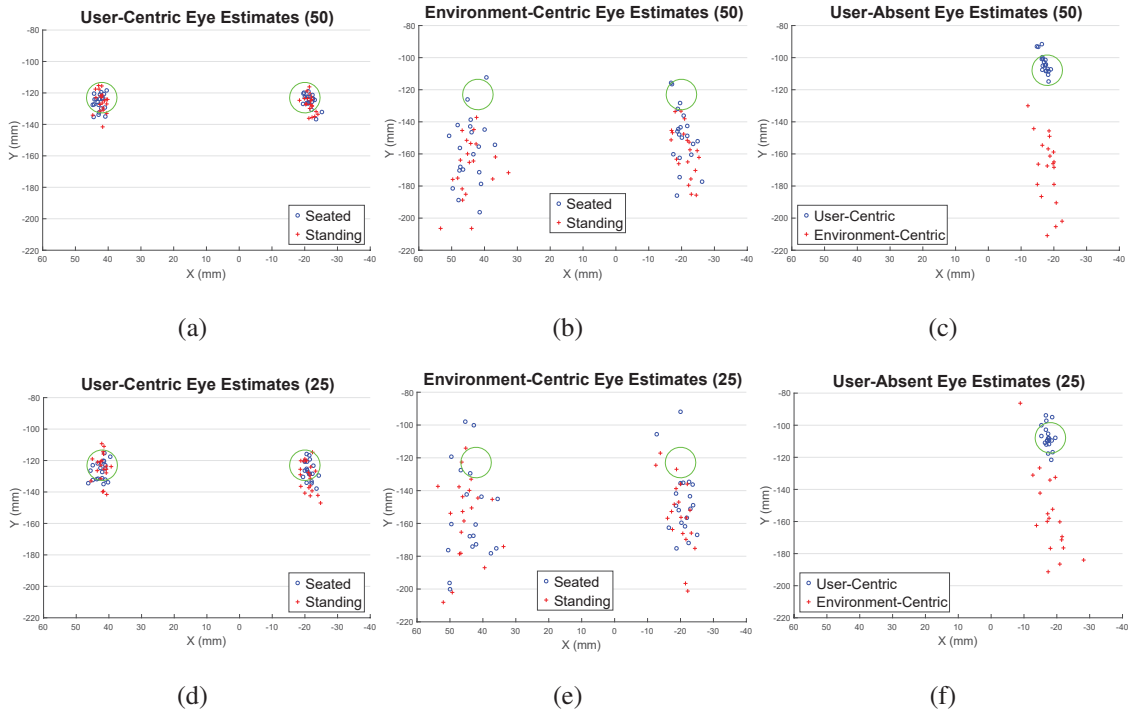


Figure 4.29 Estimated 2D user eye locations relative to the HMD marker constellation

(a) User-Centric, (b) Environment-Centric, and (c) User-Absent eye estimates after 50 Alignments. (d) User-Centric, (e) Environment-Centric, and (f) User-Absent eye estimates after 25 Alignments. In all plots, the center of the tracking constellation is at location (0, 0). Seated user calibrations are displayed in blue, standing in red. Mounted user-centric calibrations are displayed in blue, with mounted environment-centric plotted in red.

A corresponding metric to the eye location estimates is the geometrical distance to the mean location within each condition cluster. Figures 4.30 and 4.31 provide plots for distance to group medians for each calibration mode after 50 and 25 completed alignments, respectively. As seen from the 3D location plots, the user-centric calibrations, for both the user present and user-absent conditions, generate significantly tighter clusterings compared

to the environment-centric alignment types. Comparing the results after 25 with those from 50 alignments shows minimal difference with increasing alignment count. Also, there is minimal difference between the seated and standing modes for both user present conditions.

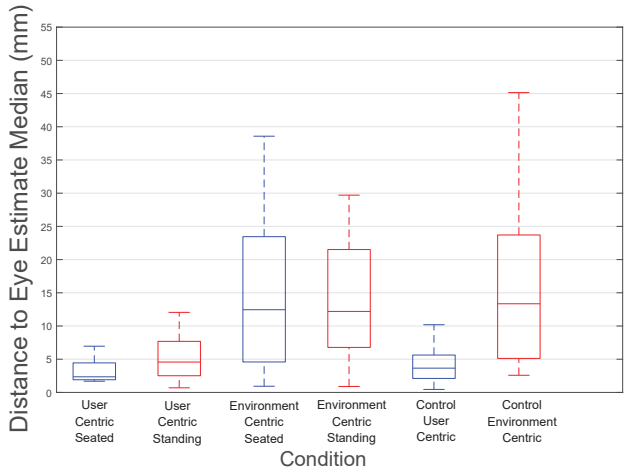


Figure 4.30 Distance to 3D eye estimate median after 50 alignments

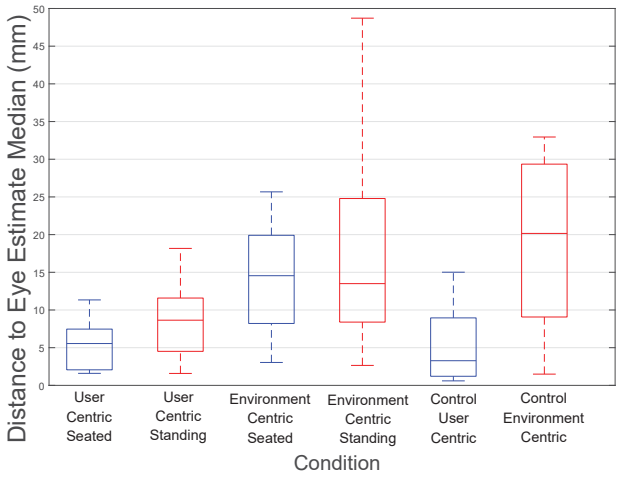


Figure 4.31 Distance to 3D eye estimate median after 25 alignments

Analysis of variance (ANOVA) statistical tests were performed in order to verify the significance, or lack there of, between conditions. The calibration number , 1–20, was used as the repeated measures variable. ANOVA between the seated and standing variants for the two user present conditions reveals no significant difference between the results at 50 alignments $(F(1,19) = 1.901, p = 0.184)$ for User-Centric and $(F(1,19) = 0.065, p = 0.802)$ for Environment-Centric alignment distances. At 25 alignments, significance is found in the User-Centric seated versus standing $(F(1,19) = 5.938, p < 0.05)$ but no significance for Environment-Centric $(F(1,19) = 0.257, p = 0.618)$. Despite the mild significance at 25 alignments for User-Centric calibrations, the remaining statistical comparisons endeavor to examine only the best case, and therefore only include the seated variants for the analysis.

As anticipated, strong significance is found between the two user-absent conditions $(F(1,19) = 14.016, p < 0.001)$ at 50 alignments and $(F(1,19) = 12.010, p < 0.01)$ at 25 alignments. Likewise, significance is found between the user-centric and environment-centric user present conditions $(F(1,19) = 17.569, p < 0.001)$ $(F(1,19) = 8.836, p < 0.01)$ at 50 and 25 alignments respectively. There is, however, no significant difference between the user present and control values for user-centric alignments $(F(1,19) = 0.833, p = 0.373)$ $(F(1,19) = 0.0, p = 0.989)$ at 50 and 25 alignments respectively. Similarly, no significant difference is reported between the user present and control values for environment-centric alignment $(F(1,19) = 0.637, p = 0.435)$ $(F(1,19) = 0.507, p = 0.485)$ at 50 and 25 alignments respectively.

Final ANOVA compares the 25 and 50 alignment values within each alignment condition itself, again only considering the seated user present modalities. There is a slight significance present between the 25 and 50 alignment values for the user-centric user present condition $(F(1,19) = 6.490, p < 0.05)$, however no further significance was found for any of the remaining conditions $(F(1,19) = 1.994, p = 0.174)$, $(F(1,19) = 0.462, p = 0.505)$, $(F(1,19) = 2.111, p = 0.163)$ for the environment-centric user present, user-centric control, and environment-centric control conditions respectively.

4.6.3.2 Reprojection Error

The reprojection error is obtained by taking the difference between the actual on-screen reticle location and its corresponding 3D target point location transformed by the projection matrix result. Figure 4.32 provides plots of the reprojection error relative to the ground truth alignment values for each condition after 25 completed alignments.

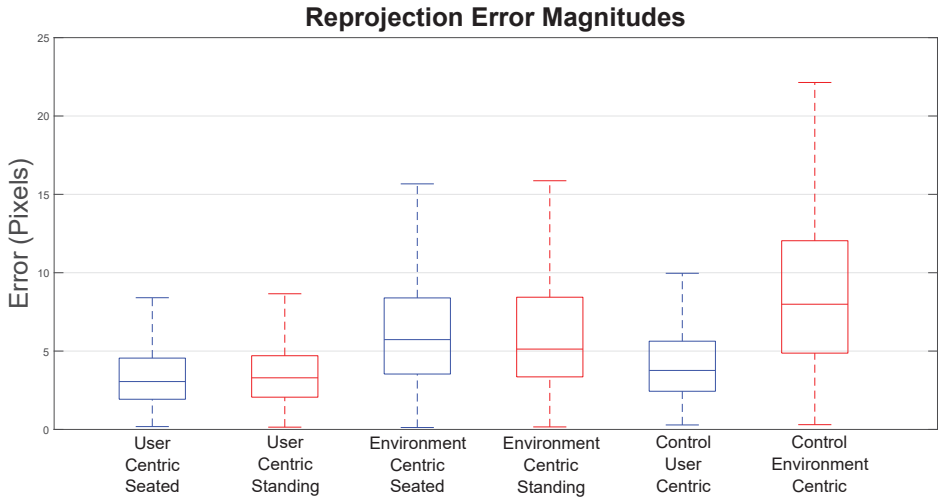


Figure 4.32 Absolute reprojection error after 25 alignments for each calibration condition

Inspection of the plot shows that the user-centric calibrations yield an expected reprojection error of less than 5 pixels, while the environment-centric modality results in slightly higher, though still less than 10, pixels in error. It is also interesting to note that the control condition does not produce any significantly lower reprojection error compared to the two user present conditions, and that the seated user does not provide any significant gains compared to a standing participant.

4.6.3.3 Results Variance Across Alignments

The final metric utilized in this analysis is a comparison of the convergence, or trend, of the calibration results with increasing alignment count. This measure indicates the threshold of alignments at which the maximum calibration gains are expected to be achieved. While it is possible to produce an alignment trend graph for every metric utilized thus far, this analysis focuses on the change in variance of the extrinsic eye location values. While Figures 4.31 and 4.30 provide these values at the 25 and 50 alignment steps, Figures 4.33, 4.34, 4.35, 4.36, 4.38, 4.37 provide the distance to median values for each condition over all 50 alignments. It is important to note though, that no results are attainable until a minimum of 6 alignments have been conducted. Since the first estimates are quite erroneous, the plots begin at alignment 9, producing 41 actual values for comparison.

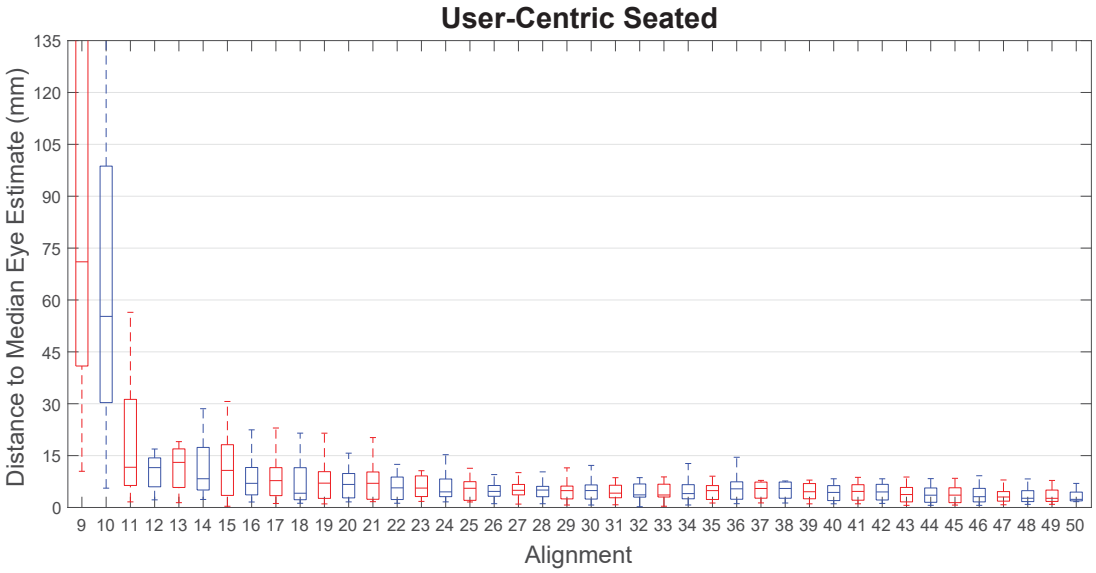


Figure 4.33 Distance to median eye estimate for the User-Centric seated condition

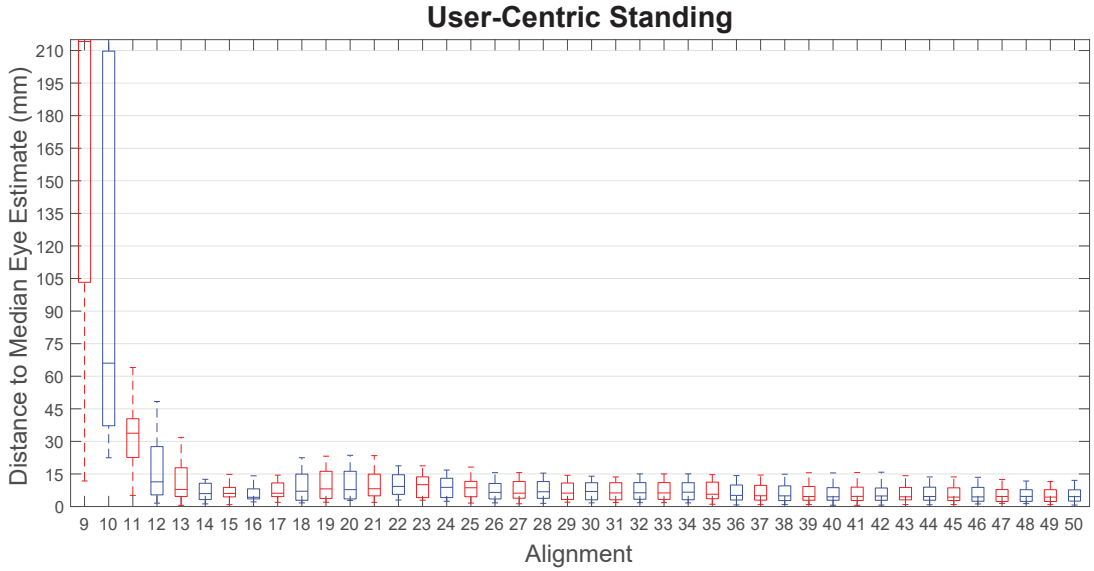


Figure 4.34 Distance to median eye estimate for the User-Centric standing condition

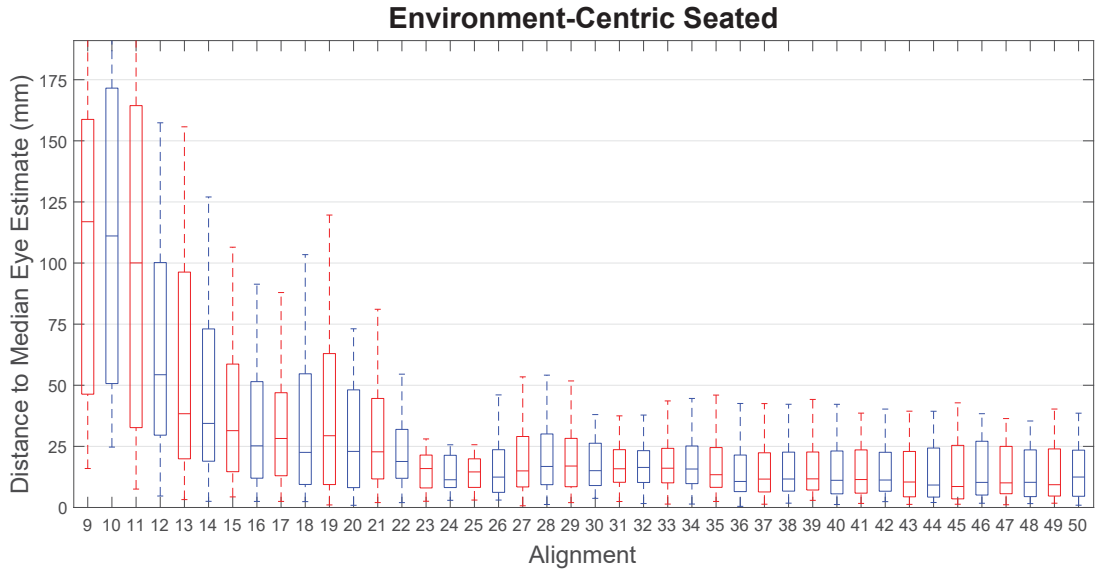


Figure 4.35 Distance to median eye estimate for Environment-Centric seated

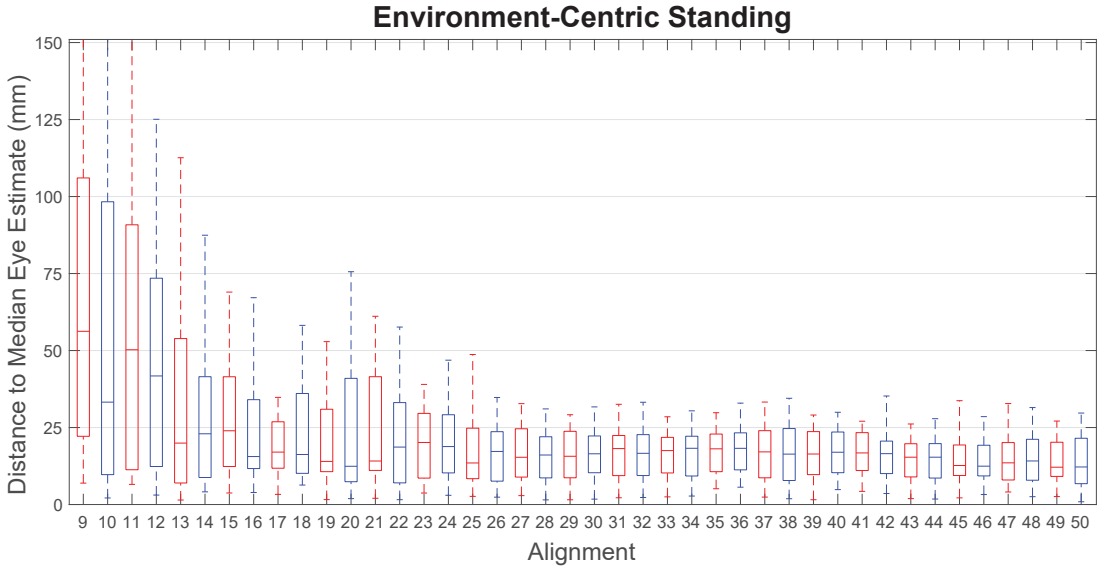


Figure 4.36 Distance to median eye estimate for Environment-Centric standing

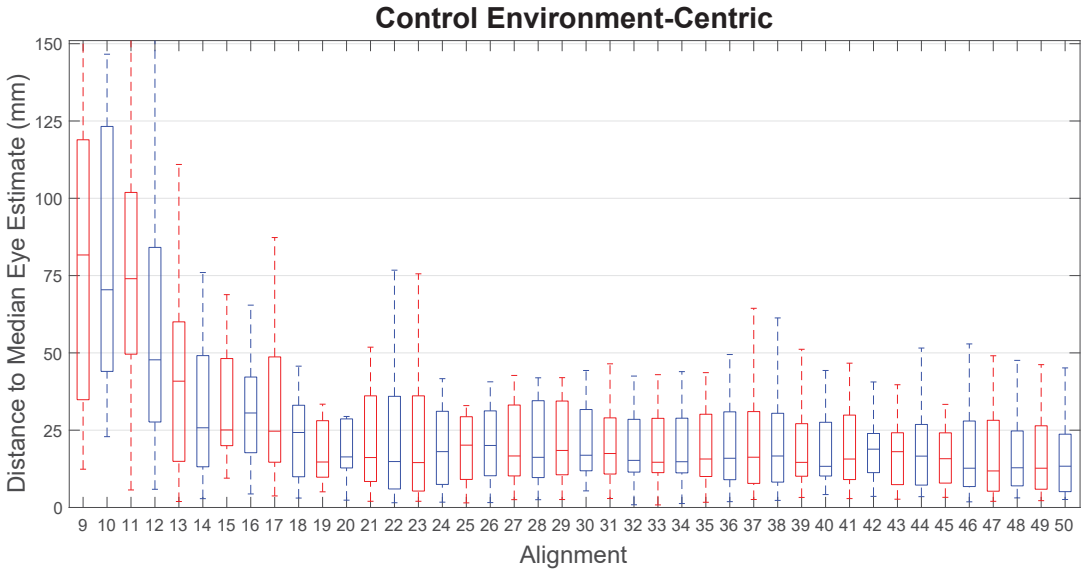


Figure 4.37 Distance to median eye estimate for control user-centric alignments

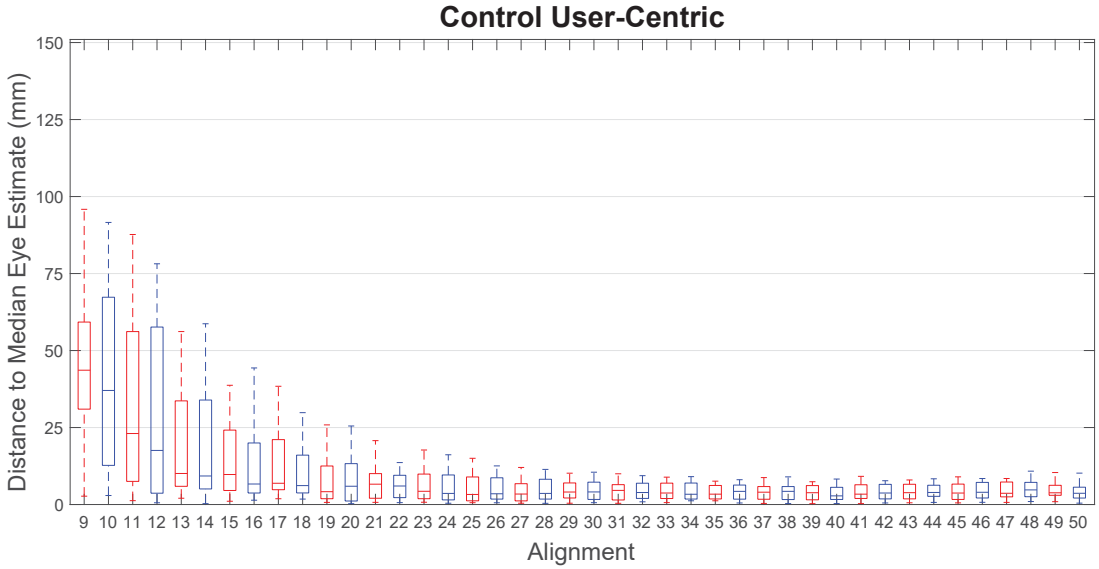


Figure 4.38 Distance to median eye estimate for control environment-centric alignments

4.6.4 Discussion and Conclusion

The objective of this experiment was multifaceted. First, the user-centric alignment design from Study 2 was repeated in order to confirm the findings for low variation in ex-

trinsic eye location estimates under calibrations utilizing arms length alignment distances. Secondly, a comparison between user-centric and environment-centric alignment modalities using an identical tracking mechanism and system setup was desired. Finally, an investigation into the actual effects of user alignment error on calibration results was facilitated through a user-absent control condition in which the HMD was affixed to an external tri-pod.

The user-centric calibration results obtained in this study match very closely to those from Study 2. This confirms that the low variance from Study 2 was not a result of the Leap Motion tracking specifically, but is indeed due to the arms length alignment distances employed during calibration. This is, again, verified, by the stark contrast to the extrinsic eye location estimates obtained from the environment-centric calibration results. The user present environment-centric eye locations, and reprojection error mirror those from prior studies, including Study 1. This evidence conclusively shows that a user-centric alignment methodology will result in much more predictable and consistent calibration outcomes compared to the more prevalent environment-centric alignment techniques. This is an especially important outcome for those researchers employing OST HMDs for registration critical tasks, and beckons an earnest consideration from the community at large for new efforts to generate standardized calibration practices applicable across device types.

It is possible, though, that the larger errors in environment-centric outcomes is a byproduct of exaggerated user alignment error, resulting from the potential for larger user misalignments due to angular movement of the head at greater distances from the alignment target. The control, user-absent, condition, however, refutes this hypothesis. Instead, the

control calibration results match nearly identically to the user present outcomes. The statistical analysis, in fact, showed no significant difference between the respective alignment modalities. This is an unexpected finding. As discussed in more thorough in Chapter 3, it has been established that alignment error is a known cause of calibration error, for manual SPAAM-like approaches. The outcomes of this work, though, reveal that the impact of these errors can be greatly ameliorated by adopting a user-centric alignment strategy. Also of note, is the lack of significance between the environment-centric results at 25 and 50 alignments. It has been suspected that user error can be, somewhat, amended through increasing alignment. Unfortunately, the examination of eye estimate variance over alignment count, and the ANOVA analysis showed no significant improvement even after doubling the calibration alignments from 25, as used in Study 2, to 50.

Even though this work clearly shows that user-centric manual calibration practices offer a clear advantage, in terms of predictability and accuracy consistency, it is still uncertain whether there is a perceptual improvement in registration accuracy. Future investigations must need to examine the subjective quality of a user-centric calibration at not only near-field but also medium and far-field environmental distances as well. It can be presumed from the outcomes of this work and Study 2, that a user-centric design will provide acceptable registration quality for AR applications intended for manufacturing or maintenance in which works are engaged in tasks at arms length. The future follow-up study should extend the visual work load to situational awareness scenarios in which the user must draw from more distant environmental markers for task completion.

CHAPTER 5

CONCLUSIONS

Augmented reality, like virtual reality, is poised to become an essential medium within our modern culture, for not only entertainment and education, but also industrial, medical, and countless other societal functions. Development of light weight low cost sensor, processing, and display hardware is speeding the delivery of several consumer priced head-mounted AR devices, including optical see-through systems, even at this present time. The increased accessibility of these devices has produced an imminent requirement for robust standardized calibration procedures which can be easily deployed and utilized by novice users.

Automatic calibration of OST HMD hardware has been shown to be a viable option through the utilization of computer vision based algorithms to localize the 6 DOF pose of the user's eye within the device at run-time. These approaches, whether employing an iris detection or corneal reflection tracking scheme, require built-in facilities to image the user's eye on-line. Unfortunately, current and upcoming HMD hardware is still absent of eye-tracking cameras and related hardware making these approaches largely inaccessible. It is expected, however, that the demand and utility of eye-tracking technologies will eventually result in cost effective hardware solutions that manufacturers will be able to easily

integrate into both VR and AR headsets alike. Until then, though, manual calibration approaches, such as the Single Point Active Alignment Method, will remain the only viable option for OST HMD calibration.

Prior research focusing on modeling and improving the accuracy of SPAAM-like procedures have produced a number of variations and provided insight into the prevailing trends for environment-centric calibration. Most notably, it has been shown that varying the distance at which screen-world alignments are taken will improve the accuracy and consistency of the results across repeated calibrations. It has also been shown though, that there is an apparent limitation on possible results accuracy, often illustrated through the large variation in extrinsic eye location estimates along the lateral , front–back, direction taken from calibration results. It has been hypothesized that this deviation may be a result of alignment inaccuracies incurred due to user error as a result of postural sway, or involuntary motor control, during the alignment process.

This work has set forth to address several of the remaining issues regarding calibration of OST HMD systems, including a comparison between the viability of automatic eye imaging methodologies compared to standard manual methods, the effect of utilizing user-centric alignment processes over the more common environment-centric, and the need for a more quantifiable metric on the effect of human alignment error on overall calibration results.

Study 1, section 4.1, is the first formal study to investigate the accuracy potential of the first automatic OST HMD calibration method, INDICA, in a direct comparison with SPAAM using novice participants within a registration critical AR task. Results from this

investigation confirm that INDICA is able to match or potentially exceed the quality attainable from SPAAM. Also of note, was the finding that a degraded calibration condition, one in which the HMD has been removed and replaced with previous calibration results re-used, did not show any significant signs of registration degradation. This is of particular benefit to system designers constructing applications where recalibration would be too tedious or cumbersome. Unfortunately, this study only utilized a monocular HMS system, and final performance within a binocular stereo system have yet to be obtained. Also, the SPAAM implementation for Study 1 utilized an environment-centric alignment process, and did not cross examine any calibration difference for user-centric manual techniques.

Study 2, section 4.2, targeted the notion of user-centric manual SPAAM calibration, and is the first full study to also investigate the utility of the Leap Motion, consumer level hand tracking device, for facilitating system agnostic calibration. This study directly examined two modalities of manual alignment, the first utilizing the participant's own hand and finger, and the second utilizing a simple stylus-like tool. Since an alignment with a stylus to an on-screen reticle is far less ambiguous than aligning to a point on a finger-tip, additional reticle types were used for finger alignment calibration sets. The results of this study show that both calibration types, finger or stylus, actually produced calibration results with more accuracy and far less variation than that seen in prior studies. Additionally, the use of more contextual reticles improved the robustness of finger alignments, though not nearly to the degree of accuracy seen for the stylus alignments. Study 2 is particularly beneficial to current OST AR developers, showing that user-centric approaches are viable

and may be the preferred method of choice given the increasing prevalence of hand and finger tracking sensors on modern HMD devices.

Study 3, section 4.3, expands on the notion of user-centric calibration and utilizes the methodology to construct a ubiquitous calibration approach for system agnostic calibration of an OST HMD. The Leap Motion controller is again utilized for the actual SPAAM calibration process. Using this method, it is possible for users to calibrate the HMD to the Leap Motion, and then utilize a secondary tracking mechanism, such as an outside-in optical IR tracking camera pair, to then facilitate 6 DOF immerssive interaction within an AR environment. In order to accomplish this, a novel calibration approach was developed to allow the determination of the transformation between the Leap Motion and secondary tracking coordinate frame using the tool tracking capabilities of the Leap Motion. Through the use of simply constructed physical jigs, it is possible to record correspondence points to reference points in both frames of reference, then through a standard absolute-orientation calculation, the final transformation obtained. This approach allows the same HMD calibration to be re-used within any AR tracking system without the need for any further adjustments from the user themselves.

Study 4, section 4.4, offers an alternative alignment process for environment-centric manual calibration, for those instances and systems where user-centric methods may not be viable. This study implements a nonius reticle style that leverages the stereopsis present within binocular HMDs to allow a user to perform stereo SPAAM calibration without the need for any prior knowledge about the needed separation of on-screen reticles to produce 3D visual cues. The nonius reticle itself is actually two halves of a single reticle split

over each eye, so that when fused, it appears to the user to become a single reticle. The participant is then able to manually adjust the on-screen locations of the reticle halves until they fuse into a single on-screen target at the physical target point's location. This process is far more intuitive then other approaches requiring the user to fuse pre-placed on-screen reticles. A cursory analysis within the study shows that the performance is comparable and potentially superior to alternatively performing two sequential monocular calibrations, one for each eye in series.

Study 5, section 4.5, addresses an alternative evaluation approach for examining the quality and state of an OST HMD calibration. Since only the users themselves are able to actually see the quality of the registrations within the system, researchers and investigators often rely on purely objective measures, such as extrinsic eye location estimates and reprojection error, to gauge the efficacy of a calibration. This study proposes the use of frustum visualization to provide an out-side observer a view of not only the user's calibration but also a possible look through the HMD from the participant's eye point as well. This method utilizes a secondary view point within the global tracking frame through which the outside observer can visualize the projection frustum resulting from a SPAAM calibration overlaid onto a user. By including an additional frame buffer to the rendering, the participant's view through the HMD can also be added to mimic the system's imaging plane. Extension of this same technique can be easily made to provide a simulated direct view through the system, if a known model of the user's environment is known, or able to be created through depth sensors, at run-time. Additionally, tele-presence and remote collaboration AR sys-

tems would also benefit from this approach since it would allow both users an option to view a task from either's vantage point.

Study 6, section 4.6, is the final study in this dissertation and expands on the findings from Study 1, 2, and 3. This investigation directly compares the potential accuracy of environment-centric alignment against a user-centric method for the same OST HMD system. A control condition, in which the user is replaced by a mechanical tri-pod system is also utilized to quantify the impact of human alignment error due to postural motion on calibration results. The outcomes of this study not only confirm those from Study 2, but conclusively verify that user-centric calibration processes produce more consistent and accurate results over the more commonly employed environment-centric strategies. The control condition also reveals that there is no significant performance degradation due to user alignment error in either alignment strategy. Additionally, no statistical significance was found in the the final calibration results between calibrations utilizing 25 or 50 alignment points. These findings, again, point researchers and developers to the use of user-centric manual calibration strategies and show that a maximum of 25 alignments is suitable.

While the extrinsic eye location estimates and reprojection errors for user-centric methodologies were significantly better than those from environment-centric processes, it is still yet to be determined if the registration quality of such a calibration would sustain viability for tasks utilizing imagery projected at medium and far visual field distances. Future research investigations are still needed to produce these subjective measures, and more precise subjective testing schemes must also be developed, in order to provide measures comparable across the growing number of devices.

Until fully automatic calibration approaches are integrated into consumer hardware, manual strategies must be deployed. It is the hope of the author that the results of this study will continue to encourage further research into the development of easily implemented system agnostic strategies of calibration, developed with novice users in mind. The development of standardized calibration practices will be largely dependent on the acceptance of the forth coming HMD options, and it will be the responsibility of the AR community at large to support a consensus on applicable practices and approaches for application developers to build on. The future of AR is very bright, and its impact on our society may be the greatest of any technology to date. It will be extremely exciting to see where future generations will take this medium in the years to come.

REFERENCES

[1] R. S. Allison, L. R. Harris, M. Jenkin, U. Jasiobedzka, and J. E. Zacher, "Tolerance of temporal delay in virtual environments," *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, 2001, pp. 247–254.

[2] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 5, 1987, pp. 698–700.

[3] M. Axholt, "Pinhole camera calibration in the presence of human noise," *Doctoral thesis*, 2011.

[4] M. Axholt, S. Peterson, and S. R. Ellis, "User boresight calibration precision for large-format head-up displays," *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. ACM, 2008, pp. 141–148.

[5] M. Axholt, S. Peterson, and S. R. Ellis, "User boresighting for ar calibration: A preliminary analysis," *Virtual Reality Conference, 2008. VR'08. IEEE*. IEEE, 2008, pp. 43–46.

[6] M. Axholt, S. D. Peterson, and S. R. Ellis, "Visual Alignment Accuracy in Head Mounted Optical See-Through AR Displays: Distribution of Head Orientation Noise," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 2009, vol. 53, pp. 2024–2028.

[7] M. Axholt, S. D. Peterson, and S. R. Ellis, "Visual alignment precision in optical see-through ar displays: Implications for potential accuracy," *Proceedings of the ACM/IEEE Virtual Reality International Conference*, 2009.

[8] M. Axholt, M. Skoglund, S. D. Peterson, M. D. Cooper, T. B. Schön, F. Gustafsson, A. Ynnerman, and S. R. Ellis, "Optical see-through head mounted display direct linear transformation calibration robustness in the presence of user alignment noise," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 2010, vol. 54, pp. 2427–2431.

[9] M. Axholt, M. A. Skoglund, S. D. OConnell, M. D. Cooper, S. R. Ellis, and A. Ynnerman, "Accuracy of Eyepoint Estimation in Optical See-Through Head-Mounted Displays Using the Single Point Active Alignment Method," *IEEE Virtual Reality Conference 2012, Orange County (CA), USA*, 2011.

[10] M. Axholt, M. A. Skoglund, S. D. O'Connell, M. D. Cooper, S. R. Ellis, and A. Ynnerman, "Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration," *Virtual Reality Conference (VR), 2011 IEEE*. IEEE, 2011, pp. 27–34.

[11] R. Azuma, "Tracking requirements for augmented reality," *Communications of the ACM*, vol. 36, no. 7, 1993, pp. 50–51.

[12] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *Computer Graphics and Applications, IEEE*, vol. 21, no. 6, 2001, pp. 34–47.

[13] R. T. Azuma, *Predictive tracking for augmented reality*, doctoral dissertation, Citeseer, 1995.

[14] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and virtual environments*, vol. 6, no. 4, 1997, pp. 355–385.

[15] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," *ACM SIGGRAPH Computer Graphics*. ACM, 1992, vol. 26, pp. 203–210.

[16] W. Barfield, *Fundamentals of wearable computers and augmented reality*, CRC Press, 2015.

[17] W. Barfield and E. Danas, "Comments on the use of olfactory displays for virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 1, 1996, pp. 109–121.

[18] W. Barfield and T. A. Furness, *Virtual environments and advanced interface design*, Oxford University Press, 1995.

[19] G. Beach, C. J. Cohen, J. Braun, and G. Moody, "Eye tracker system for use with head mounted displays," *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*. IEEE, 1998, vol. 5, pp. 4348–4352.

[20] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive group-to-group telepresence," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 4, 2013, pp. 616–625.

[21] A. H. Behzadan and V. R. Kamat, "Georeferenced registration of construction graphics in mobile outdoor augmented reality," *Journal of Computing in Civil Engineering*, vol. 21, no. 4, 2007, pp. 247–258.

[22] M. Billinghurst, H. Kato, and I. Poupyrev, "The magicbook-moving seamlessly between reality and virtuality," *Computer Graphics and Applications, IEEE*, vol. 21, no. 3, 2001, pp. 6–8.

[23] O. Bimber and R. Raskar, "Modern approaches to augmented reality," *ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 1.

[24] F. A. Biocca and J. P. Rolland, "Virtual eyes can rearrange your body: Adaptation to visual displacement in see-through, head-mounted displays," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, 1998, pp. 262–277.

[25] J. Blake and H. B. Gurocak, "Haptic glove with MR brakes for virtual reality," *Mechatronics, IEEE/ASME Transactions on*, vol. 14, no. 5, 2009, pp. 606–615.

[26] M. Bonacker, E. Schubert-Alshuth, and W. Jaschinski, "Precise placement of nonius lines on a personal computer screen for measuring fixation disparity," *Ophthalmic and Physiological Optics*, vol. 14, no. 3, 1994, pp. 317–319.

[27] E. Bostanci, N. Kanwal, S. Ehsan, and A. F. Clark, "Tracking methods for augmented reality," *The 3rd international conference on machine vision*, 2010, pp. 425–429.

[28] D. A. Bowman and R. P. McMahan, "Virtual reality: how much immersion is enough?," *Computer*, vol. 40, no. 7, 2007, pp. 36–43.

[29] G. Burdea and P. Coiffet, "Virtual reality technology," *Presence: Teleoperators and virtual environments*, vol. 12, no. 6, 2003, pp. 663–664.

[30] O. Cakmakci and J. Rolland, "Head-worn displays: a review," *Display Technology, Journal of*, vol. 2, no. 3, 2006, pp. 199–216.

[31] M. Camplani and L. Salgado, "Efficient spatio-temporal hole filling strategy for kinect depth maps," *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82900E–82900E.

[32] T. P. Caudell and D. W. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*. IEEE, 1992, vol. 2, pp. 659–669.

[33] T. R. Coles, N. W. John, D. A. Gould, and D. G. Caldwell, "Integrating haptics with augmented reality in a femoral palpation and needle insertion training simulation," *Haptics, IEEE Transactions on*, vol. 4, no. 3, 2011, pp. 199–209.

[34] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy, "Augmented touch without visual obtrusion," *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 99–102.

[35] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 1993, pp. 135–142.

[36] M. Czernuszenko, D. Pape, D. Sandin, T. DeFanti, G. L. Dawe, and M. D. Brown, "The ImmersaDesk and Infinity Wall projection-based virtual reality displays," *ACM SIGGRAPH Computer Graphics*, vol. 31, no. 2, 1997, pp. 46–49.

[37] M. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *Robotics and Automation, IEEE Transactions on*, vol. 17, no. 3, 2001, pp. 229–241.

[38] D. Dobler, M. Haller, and P. Stampfl, "ASR: augmented sound reality," *ACM SIGGRAPH 2002 conference abstracts and applications*. ACM, 2002, pp. 148–148.

[39] D. Drascic and P. Milgram, "Perceptual issues in augmented reality," *Electronic Imaging: Science & Technology*. International Society for Optics and Photonics, 1996, pp. 123–134.

[40] U. Eck and C. Sandor, *HARP: A framework for visuo-haptic augmented reality*, IEEE, 2013.

[41] S. R. Ellis, M. J. Young, B. D. Adelstein, and S. M. Ehrlich, "Discrimination of changes of latency during voluntary hand movement of virtual objects," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1999, vol. 43, pp. 1182–1186.

[42] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," *Computer Vision–ECCV 2014*, Springer, 2014, pp. 834–849.

[43] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT press, 1993.

[44] S. Feiner, B. Macintyre, and D. Seligmann, "Knowledge-based augmented reality," *Communications of the ACM*, vol. 36, no. 7, 1993, pp. 53–62.

[45] E. Foxlin, "Handbook of Virtual Environment Technologies, chapter Motion Tracking Technologies and Requirements," 2002.

[46] Y. Furukawa and J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[47] Y. Genc, M. Tuceryan, and N. Navab, "Practical solutions for calibration of optical see-through devices," *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2002, p. 169.

[48] J. Grubert, J. Tuemle, R. Mecke, and M. Schenk, "Comparative User Study of two See-through Calibration Methods.," *VR*, vol. 10, 2010, pp. 269–270.

[49] T. Gunther, I. S. Franke, and R. Groh, "Aughanded Virtuality-the hands in the virtual environment," *3D User Interfaces (3DUI), 2015 IEEE Symposium on*. IEEE, 2015, pp. 157–158.

[50] M. Harders, G. Bianchi, B. Knoerlein, and G. Székely, "Calibration, registration, and synchronization for high precision augmented reality haptics," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 1, 2009, pp. 138–149.

[51] R. Hartley and S. B. Kang, "Parameter-free radial distortion correction with center of distortion estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 8, 2007, pp. 1309–1321.

[52] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Robotica*, vol. 23, no. 2, 2005, pp. 271–271.

[53] S. Henderson and S. Feiner, "Exploring the benefits of augmented reality documentation for maintenance and repair," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 10, 2011, pp. 1355–1368.

[54] S. J. Henderson and S. Feiner, "Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret," *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 135–144.

[55] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, 2012, pp. 647–663.

[56] A. Henrysson and M. Ollila, "UMAR: Ubiquitous mobile augmented reality," *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*. ACM, 2004, pp. 41–45.

[57] J. Herling and W. Broll, "Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments," *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, pp. 207–212.

[58] J. D. Hol, T. B. Schön, F. Gustafsson, and P. J. Slycke, "Sensor fusion for augmented reality," *Information Fusion, 2006 9th International Conference on*. IEEE, 2006, pp. 1–6.

[59] T. Höllerer, S. Feiner, D. Hallaway, B. Bell, M. Lanzagorta, D. Brown, S. Julier, Y. Baillot, and L. Rosenblum, "User interface management techniques for collaborative mobile augmented reality," *Computers & Graphics*, vol. 25, no. 5, 2001, pp. 799–810.

[60] R. L. Holloway, "Registration error analysis for augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, 1997, pp. 413–432.

[61] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, 1987, pp. 629–642.

[62] M. Huber, D. Pustka, P. Keitler, F. Echtler, and G. Klinker, "A system architecture for ubiquitous tracking environments," *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2007, pp. 1–4.

[63] J. F. Hughes, A. Van Dam, J. D. Foley, and S. K. Feiner, *Computer graphics: principles and practice*, Pearson Education, 2013.

[64] Y. Itoh and G. Klinker, "Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization," *3D User Interfaces (3DUI), 2014 IEEE Symposium on*. IEEE, 2014, pp. 75–82.

[65] Y. Itoh and G. Klinker, "Performance and sensitivity analysis of indica: Interaction-free display calibration for optical see-through head-mounted displays," *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 171–176.

[66] Y. Itoh and G. Klinker, "Light-field correction for spatial calibration of optical see-through head-mounted displays," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 21, no. 4, 2015, pp. 471–480.

[67] Y. Itoh, F. Pankrat, C. Waechter, and G. Klinker, "Calibration of Head-Mounted Finger Tracking to Optical See-Through Head Mounted Display,", Demonstration at 12th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013, Adelaide, Australia, October 1-4, 2013, 2013.

[68] H. Iwata, "Walking about virtual environments on an infinite floor," *Virtual Reality, 1999. Proceedings., IEEE*. IEEE, 1999, pp. 286–293.

[69] M. C. Jacobs, M. A. Livingston, et al., "Managing latency in complex augmented reality systems," *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM, 1997, pp. 49–ff.

[70] S. Jeon and S. Choi, "Haptic augmented reality: Taxonomy and an example of stiffness modulation," *Presence: Teleoperators and Virtual Environments*, vol. 18, no. 5, 2009, pp. 387–408.

[71] M. Kassner, W. Patera, and A. Bulling, "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction," *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 1151–1160.

[72] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*. IEEE, 1999, pp. 85–94.

[73] B. F. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais, "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, vol. 16, no. 4, 2012, pp. 253–269.

[74] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007, pp. 225–234.

[75] G. Klein and D. Murray, "Parallel tracking and mapping on a camera phone," *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 83–86.

[76] G. S. Klein and T. W. Drummond, "Tightly integrated sensor fusion for robust visual tracking," *Image and Vision Computing*, vol. 22, no. 10, 2004, pp. 769–776.

[77] G. Klinker, D. Stricker, and D. Reiners, "Augmented Reality: A Balance Act between High Quality and Real-Time Constraints," *Mixed Reality–Merging Real and Virtual Worlds, Ohmsha & Springer Verlag*, 1999, pp. 325–346.

[78] B. Kress and T. Starner, "A review of head-mounted displays (HMD) technologies and applications for consumer electronics," *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87200A–87200A.

[79] T. Langlotz, S. Mooslechner, S. Zollmann, C. Degendorfer, G. Reitmayr, and D. Schmalstieg, "Sketching up the world: in situ authoring for mobile Augmented Reality," *Personal and ubiquitous computing*, vol. 16, no. 6, 2012, pp. 623–630.

[80] A. Lécuyer, J.-M. Burkhardt, and L. Etienne, "Feeling bumps and holes without a haptic interface: the perception of pseudo-haptic textures," *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004, pp. 239–246.

[81] M. A. Livingston, L. J. Rosenblum, S. J. Julier, D. Brown, Y. Baillot, I. Swan, J. L. Gabbard, D. Hix, et al., *An augmented reality system for military operations in urban terrain*, Tech. Rep., DTIC Document, 2002.

[82] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-d vision: from images to geometric models*, vol. 26, Springer Science & Business Media, 2012.

[83] W. E. Mackay, "Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers," *Proceedings of the working conference on Advanced visual interfaces*. ACM, 1998, pp. 13–21.

[84] P. Maier, A. Dey, C. A. Waechter, C. Sandor, M. Tönnis, and G. Klinker, "An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration," *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE, 2011, pp. 267–268.

[85] W. R. Mark, L. McMillan, and G. Bishop, "Post-rendering 3D warping," *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM, 1997, pp. 7–ff.

[86] T. Mazuryk and M. Gervautz, "Virtual reality-history, applications, technology and future," 1996.

[87] E. McGarrity, Y. Genc, M. Tuceryan, C. Owen, and N. Navab, "A new system for online quantitative evaluation of optical see-through augmentation," *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*. IEEE, 2001, pp. 157–166.

[88] S. P. McKee and D. M. Levi, "Dichoptic hyperacuity: the precision of nonius alignment," *JOSA A*, vol. 4, no. 6, 1987, pp. 1104–1108.

[89] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks Jr, "Effect of latency on presence in stressful virtual environments," *virtual reality, 2003. Proceedings. IEEE*. IEEE, 2003, pp. 141–148.

[90] S. Meerits and H. Saito, "Real-Time Diminished Reality for Dynamic Scenes," *Mixed and Augmented Reality Workshops (ISMARW), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 53–59.

[91] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, 1994, pp. 1321–1329.

[92] M. Montemerlo and S. Thrun, "Simultaneous localization and mapping with unknown data association using FastSLAM," *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*. IEEE, 2003, vol. 2, pp. 1985–1991.

[93] K. Moser, Y. Itoh, K. Oshima, J. E. Swan, G. Klinker, and C. Sandor, "Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 21, no. 4, 2015, pp. 491–500.

[94] K. R. Moser, S. Anreddy, and J. E. Swan II, "Calibration and Interaction in Optical See-Through Augmented Reality using Leap Motion," *Virtual Reality (VR), 2016 IEEE*. IEEE, 2016.

[95] K. R. Moser, S. Anreddy, and J. E. Swan II, "Leap Motion Hand and Stylus Tracking for Calibration and Interaction within Optical See-Through Augmented Reality," *Virtual Reality (VR), 2016 IEEE*. IEEE, 2016.

[96] K. R. Moser and J. E. Swan, "Evaluating optical see-through head-mounted display calibration via frustum visualization," *Virtual Reality (VR), 2015 IEEE*. IEEE, 2015, pp. 371–371.

[97] K. R. Moser and J. E. Swan, "[POSTER] Improved SPAAM Robustness through Stereo Calibration," *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 200–201.

[98] K. R. Moser and J. E. Swan, "Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller," *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2016, pp. 159–167.

[99] K. R. Moser and J. E. Swan II, "Evaluation of User-Centric Optical See-Through Head-Mounted Display Calibration Using a Leap Motion Controller," *Virtual Reality (VR), 2016 IEEE*. IEEE, 2016.

[100] R. Mur-Artal, J. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *Robotics, IEEE Transactions on*, vol. 31, no. 5, 2015, pp. 1147–1163.

[101] M. Nabiyouni, A. Saktheeswaran, D. A. Bowman, and A. Karanth, "Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality," *3D User Interfaces (3DUI), 2015 IEEE Symposium on*. IEEE, 2015, pp. 3–10.

[102] D. Nahon, G. Subileau, and B. Capel, "Never Blind VR enhancing the virtual reality headset experience with augmented virtuality," *Virtual Reality (VR), 2015 IEEE*. IEEE, 2015, pp. 347–348.

[103] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose, "Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction," *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2011, pp. 93–102.

[104] N. Navab, S. Zokai, Y. Genc, and E. M. Coelho, "An On-line Evaluation System for Optical See-through Augmented Reality," *Proceedings of the IEEE Virtual Reality 2004*. IEEE Computer Society, 2004, p. 245.

[105] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*. IEEE, 2011, pp. 127–136.

[106] T. Ni, G. S. Schmidt, O. G. Staadt, M. A. Livingston, R. Ball, and R. May, "A survey of large high-resolution display technologies, techniques, and applications," *Virtual Reality Conference, 2006*. IEEE, 2006, pp. 223–236.

[107] C. Nitschke, A. Nakazawa, and H. Takemura, "Corneal imaging revisited: An overview of corneal reflection analysis and applications," *Information and Media Technologies*, vol. 8, no. 2, 2013, pp. 389–406.

[108] M. O'Loughlin and C. Sandor, "User-Centric Calibration for Optical See-Through Augmented Reality," *Master thesis*, 2013.

[109] J. Orlosky, T. Toyama, K. Kiyokawa, and D. Sonntag, "ModulAR: Eye-controlled Vision Augmentations for Head Mounted Displays," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 21, no. 11, 2015, pp. 1259–1268.

[110] C. B. Owen, J. Zhou, A. Tang, and F. Xiao, "Display-relative calibration for optical see-through head-mounted displays," *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*. IEEE, 2004, pp. 70–78.

[111] M. Park, S. Serefoglou, L. Schmidt, K. Radermacher, C. Schlick, and H. Luczak, "Hand-eye coordination using a video see-through augmented reality system," *The ergonomics open journal*, vol. 1, no. 1, 2008.

[112] H. E. Pence, "Smartphones, smart objects, and augmented reality," *The Reference Librarian*, vol. 52, no. 1-2, 2010, pp. 136–145.

[113] W. Piekarski, B. Avery, B. H. Thomas, and P. Malbezin, "Hybrid indoor and outdoor tracking for mobile 3d mixed reality," *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2003, p. 266.

[114] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, "Corneal-imaging calibration for optical see-through head-mounted displays," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 21, no. 4, 2015, pp. 481–490.

[115] J. Ponce, D. Forsyth, E.-p. Willow, S. Antipolis-Méditerranée, R. dactivité RAweb, L. Inria, and I. Alumni, "Computer vision: a modern approach," *Computer*, vol. 16, no. 11, 2011.

[116] M. Qiu and S. De Ma, "The nonparametric approach for camera calibration," *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE, 1995, pp. 224–229.

[117] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller, "Using augmented virtuality for remote collaboration," *Presence: Teleoperators and virtual environments*, vol. 13, no. 3, 2004, pp. 338–354.

[118] H. Regenbrecht, C. Ott, M. Wagner, T. Lum, P. Kohler, W. Wilke, and E. Mueller, "An augmented virtuality approach to 3D videoconferencing," *Proceedings of the 2nd IEEE/ACM international Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2003, p. 290.

[119] G. Reitmayr and T. Drummond, "Going out: robust model-based tracking for outdoor augmented reality," *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2006, pp. 109–118.

[120] W. Richards, "Anomalous stereoscopic depth perception," *JOSA*, vol. 61, no. 3, 1971, pp. 410–414.

[121] J. P. Rolland, "Wide-angle, off-axis, see-through head-mounted display," *Optical engineering*, vol. 39, no. 7, 2000, pp. 1760–1767.

[122] K. Satoh, M. Anabuki, H. Yamamoto, and H. Tamura, "A hybrid registration method for outdoor augmented reality," *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*. IEEE, 2001, pp. 67–76.

[123] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof, "Global pose estimation using multi-sensor fusion for outdoor augmented reality," *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 153–162.

[124] M. J. Schuemie, P. Van Der Straaten, M. Krijn, and C. A. Van Der Mast, "Research on presence in virtual reality: A survey," *CyberPsychology & Behavior*, vol. 4, no. 2, 2001, pp. 183–201.

[125] B. Schwald and B. De Laval, "An augmented reality system for training and assistance to maintenance in the industrial context," 2003.

[126] S. Shah and J. Aggarwal, "Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation*," *Pattern Recognition*, vol. 29, no. 11, 1996, pp. 1775–1788.

[127] K. Shimono, H. Ono, S. Saida, and A. P. Mapp, "Methodological caveats for monitoring binocular eye position with nonius stimuli," *Vision research*, vol. 38, no. 4, 1998, pp. 591–600.

[128] K. T. Simsarian and K.-P. Akesson, "Windows on the world: An example of augmented virtuality," 1997.

[129] M. Singh and M. P. Singh, "Augmented reality interfaces," *IEEE Internet Computing*, vol. 6, 2013, pp. 66–70.

[130] F. Smit, R. Van Liere, S. Beck, and B. Fröhlich, "An image-warping architecture for vr: Low latency versus image quality," *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 2009, pp. 27–34.

[131] J. Sodnik, S. Tomazic, R. Grasset, A. Duenser, and M. Billinghurst, "Spatial sound localization in an augmented reality environment," *Proceedings of the 18th Australia conference on computer-human interaction: design: activities, artefacts and environments*. ACM, 2006, pp. 111–118.

[132] D. A. Southard, "Viewing model for stereoscopic head-mounted displays," *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1994, pp. 119–129.

[133] M. A. Srinivasan and C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges," *Computers & Graphics*, vol. 21, no. 4, 1997, pp. 393–404.

[134] K. M. Stanney, R. S. Kennedy, J. M. Drexler, and D. L. Harm, "Motion sickness and proprioceptive aftereffects following virtual environment exposure," *Applied Ergonomics*, vol. 30, no. 1, 1999, pp. 27–38.

[135] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence*, vol. 7, no. 4, 1998, pp. 327–351.

[136] R. J. Stone, "Haptic feedback: A brief history from telepresence to virtual reality," *Haptic Human-Computer Interaction*, Springer, 2001, pp. 1–16.

[137] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto, "Camera models and fundamental concepts used in geometric computer vision," *Foundations and Trends R in Computer Graphics and Vision*, vol. 6, no. 1–2, 2011, pp. 1–183.

[138] I. E. Sutherland, "The ultimate display," *Multimedia: From Wagner to virtual reality*, 1965.

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

[140] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012, pp. 173–176.

[141] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, pp. 73–80.

[142] A. Tang, J. Zhou, and C. Owen, "Evaluation of calibration procedures for optical see-through head-mounted displays," *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2003, p. 161.

[143] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and S. I. MacKenzie, "Effects of tracking technology, latency, and spatial jitter on object movement," *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*. IEEE, 2009, pp. 43–50.

[144] J. N. Templeman, P. S. Denbrook, and L. E. Sibert, "Virtual locomotion: Walking in place through virtual environments," *Presence: teleoperators and virtual environments*, vol. 8, no. 6, 1999, pp. 598–617.

[145] N. A. Thacker and J. E. Mayhew, "Optimal combination of stereo camera calibration from arbitrary stereo images," *Image and vision computing*, vol. 9, no. 1, 1991, pp. 27–32.

[146] S. Thrun and J. J. Leonard, "Simultaneous localization and mapping," *Springer handbook of robotics*, Springer, 2008, pp. 871–889.

[147] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *Robotics and Automation, IEEE Journal of*, vol. 3, no. 4, 1987, pp. 323–344.

[148] M. Tuceryan and N. Navab, "Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR," *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*. IEEE, 2000, pp. 149–158.

[149] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 4, 1991, pp. 376–380.

[150] A. Utsumi, P. Milgram, H. Takemura, and F. Kishino, "Investigation of errors in perception of stereoscopically presented virtual object locations in real display space," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1994, vol. 38, pp. 250–254.

[151] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *International Journal of Virtual Reality*, vol. 9, no. 2, 2010, p. 1.

[152] K. Vasylevska, I. Podkosova, and H. Kaufmann, "Walking in Virtual Reality: Flexible Spaces and Other Techniques," *The Visual Language of Technique*, Springer, 2015, pp. 81–97.

[153] J. Vince, *Mathematics for computer graphics*, Springer Science & Business Media, 2013.

[154] C. Y. Vincent and T. Tjahjadi, "Multiview camera-calibration framework for non-parametric distortions removal," *IEEE Transactions on Robotics*, vol. 21, no. 5, 2005, pp. 1004–1009.

[155] D. Wagner, T. Pintaric, F. Ledermann, and D. Schmalstieg, *Towards massively multi-user augmented reality on handheld devices*, Springer, 2005.

[156] P. Wellner, W. Mackay, and R. Gold, "Computer-augmented environments: back to the real world," *Communications of the ACM*, vol. 36, no. 7, 1993, pp. 24–27.

[157] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 10, 1992, pp. 965–980.

[158] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr, "Effects of stereo viewing conditions on distance perception in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 17, no. 1, 2008, pp. 91–101.

[159] T. Yamada, S. Yokoyama, T. Tanikawa, K. Hirota, and M. Hirose, "Wearable olfactory display: Using odor in outdoor environment," *Virtual Reality Conference, 2006*. IEEE, 2006, pp. 199–206.

[160] S. Yamazaki, M. Mochimaru, and T. Kanade, "Simultaneous self-calibration of a projector and a camera using structured light," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 60–67.

[161] Y. Yanagida, S. Kawato, H. Noma, A. Tomono, and N. Tesutani, "Projection based olfactory display with nose tracking," *Virtual Reality, 2004. Proceedings. IEEE*. IEEE, 2004, pp. 43–50.

[162] S. J. Yohan, S. Julier, Y. Baillot, M. Lanzagorta, D. Brown, and L. Rosenblum, "Bars: Battlefield augmented reality system," *In NATO Symposium on Information Processing Techniques for Military Systems*. Citeseer, 2000.

[163] S. You and U. Neumann, "Fusion of vision and gyro tracking for robust augmented reality registration," *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, 2001, pp. 71–78.

[164] S. You, U. Neumann, and R. Azuma, "Hybrid inertial and vision tracking for augmented reality registration," *Virtual Reality, 1999. Proceedings., IEEE*. IEEE, 1999, pp. 260–267.

[165] S. You, U. Neumann, and R. Azuma, "Orientation tracking for outdoor augmented reality registration," *Computer Graphics and Applications, IEEE*, vol. 19, no. 6, 1999, pp. 36–42.

[166] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, 2000, pp. 1330–1334.

[167] Z. Zheng, X. Liu, H. Li, and L. Xu, "Design and fabrication of an off-axis see-through head-mounted display with an x–y polynomial surface," *Applied optics*, vol. 49, no. 19, 2010, pp. 3661–3668.

[168] Z. Zhou, A. D. Cheok, X. Yang, and Y. Qiu, "An experimental study on the role of 3D sound in augmented reality environment," *Interacting with Computers*, vol. 16, no. 6, 2004, pp. 1043–1068.