

THE UNIVERSITY OF BIRMINGHAM

**Augmented Reality
and
Scene Examination**

by

Sashah James Eftekhari

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

in the

Department of Electronic, Electrical and Computer Engineering

March 2011

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Declaration of Authorship

I, Sashah James Eftekhari, declare that this thesis titled, ‘Augmented Reality and Scene Examination’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

THE UNIVERSITY OF BIRMINGHAM

Abstract

Department of Electronic, Electrical and Computer Engineering

Doctor of Philosophy

by [Sashah James Eftekhari](#)

The research presented in this thesis explores the impact of Augmented Reality on human performance, and compares this technology with Virtual Reality using a head-mounted video-feed for a variety of tasks that relate to scene examination. The motivation for the work was the question of whether Augmented Reality could provide a vehicle for training in crime scene investigation. The Augmented Reality application was developed using fiducial markers in the Windows Presentation Foundation, running on a wearable computer platform; Virtual Reality was developed using the Crytek game engine to present a photo-realistic 3D environment; and a video-feed was provided through head-mounted webcam. All media were presented through head-mounted displays of similar resolution to provide the sole source of visual information to participants in the experiments. The experiments were designed to increase the amount of mobility required to conduct the search task, i.e., from rotation in the horizontal or vertical plane through to movement around a room. In each experiment, participants were required to find objects and subsequently recall their location. In the first experiment, objects were presented with and without background details, i.e., virtual 3D objects against a black background or superimposed on the fiducial markers in the real environment. The findings suggest that even where the real world is not task-relevant, by offering cues to spatial position participants were helped to orient their own spatial frame of reference with incidental contextual information that helped the encoding of object identity (e.g., the synthetic tank is by the real plantpot). This effect was explored further by blocking stimuli into sets of 3, this resulted in enhanced performance but also hinted that an orientation-only frame of reference adds no value to recall performance unless supplemented with a visual frame of reference. Experiment 2 compared search and recall in virtual reality,

real and augmented reality presentations of a task that required participants to find and recall coloured bottles attached to a wall. Analysis of condition and location of object in terms of recall and dwell time found that whilst recall performance between conditions was indifferent, behavioural characteristics differed somewhat in the VR condition. The subtle difference in VR behaviour was confirmed by using head-tracking results to explore dwell time on each object during each trial. Analysis of these data showed that participants attempted to search the area in similar manner for AR and RL but not for VR. Experiment 3 required participants to search for coloured bottles positioned around a room. Search was conducted in either real or augmented reality conditions, and the results implied similar search times but differences in recall ability. Navigation alone implied that the task was just as effective under AR however, recall performance in the AR condition was poor compared with real life. This seemed to result from the requirement of the AR technology to focus on registration of the fiducial marker which in turn encouraged participants to pay less attention to the surrounding environment and, therefore, miss subtle elements proven to be beneficial to recall in the first experiment, such as peripheral recognition of objects and spatial registration of the surrounding environment. Thus, experiment 3 indicates that current AR technology or at least the fiducial marker system employed is not a good measure of fully mobile human performance in the real world. It is concluded that human performance is affected not merely via the medium through which the world is perceived but moreover, the constraints governing how movement in the world is controlled.

Acknowledgements

During the long journey for which this research has undergone there are many people who I have consulted and deliberated with. To all those people for whom I hold great regard I wish to extend my warmest thank you.

First and foremost thank you to my supervisor Professor Chris Baber PhD, head of The University of Birmingham HIT Lab research team and recently promoted to Head of School, EECE department, congratulations on that by the way. Thank you for your guidance and expertise, especially in those dark ages where you didn't see me for months at a time. At least you can be confident now, I was busy working away with my square glasses for eyes. Also thank you to Dr Robert J. Houghton PhD for your guidance in statistical modelling and psychometric testing.

To my love Claire Marsh. You are my all; your support, guidance and understanding is everything I could ever hope for.

My loving thanks for all your support to Farhad (Dad), Josephine (Mum), Amy, Amir, Alicia and all my family... *(here we go)* The Eftekhari(s): Fatemeh, Farshid and Omedeh, Ali and Roohangiz, Victoria, Serioja and Daniella, Jopin and Shahla, Serina, Niki and Sepideh, Nima, Ophelia, Amir, Farshad, Tina, Rona, Tiva. The Wood(s): Stan and Sylvia, Andrew and Elaine, Phil and Jane. And how could I forget my beautiful little niece Aaliyah, nephew 'little dude' Kairen and their mum Sarah.

To my close friends. Robert Guest for your expert guidance on 3D modelling and Danny Rigby for your advise on mathematical modelling.

Less not forget the ESPRC for without their funding this research would not be possible. And finally for your friendship and kindness: Michael and Katherine Watson, Thomas Chaffe, Tim Smith, Scott Newton, Dr Stuart Proctor PhD, David Rigby, Mohammad Sharkir Saleem and Jonathan Valentine.

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iv
List of Figures	ix
List of Tables	xii
1 Introduction	1
1.1 Overview	1
1.2 Problem Statement	2
1.3 Research Assumptions	3
1.4 Research Processes	3
1.5 Domain of Investigation	4
1.5.1 Synthetic Environments - Transfer of Training	4
1.5.2 Crime Scene Investigation	5
1.5.3 Case Solving Methodology	5
1.5.4 Early Work - Proposed Augmented Reality Ubiquitous Computing System Concept for Evidence Reconstruction at a Scene of Crime	8
1.5.4.1 Evidence Retrieval	9
1.5.4.2 Evidence Representation	11
1.5.4.3 Concluding Thoughts	13
1.5.5 Transfer of Training - Virtual Environments to the Real World	13
1.5.6 Discussion	16
1.6 Contributions Made	18
1.7 Thesis Structure	19
2 Reality and Virtuality - An Introduction to AR	21
2.1 Introduction	21
2.2 RV - Combining Real and Synthetic Media	24
2.2.1 Virtual Reality	24
2.2.2 Augmented Reality	25

2.2.3	Augmented Virtuality	28
2.3	Display Systems	29
2.3.1	Head-down Displays	29
2.3.2	Head Mounted Displays	31
2.4	Tracking Technology - Tracking with Hardware	33
2.4.1	Lidar	34
2.4.2	Stereo Camera	35
2.4.3	Ultrasonic	35
2.4.4	Magnetic Trackers	36
2.4.5	Inertial Trackers	37
2.4.5.1	Accelerometer	38
2.4.5.2	Gyroscope	39
2.5	Pattern Recognition - Tracking with Software	41
2.6	Discussion	42
3	Synthetically Enhanced Reality - Human Visual Perception	46
3.1	Introduction	46
3.2	Human Perception	46
3.2.1	Body-Based Movement	48
3.2.2	Depth Perception	51
3.2.2.1	Depth Perception in AR	51
3.3	Vision	52
3.3.1	Field of View	52
3.3.2	Binocular Rivalry	54
3.3.3	Eye Offset	56
3.3.4	Visual Acuity	56
3.3.4.1	Visual Acuity in AR	57
3.3.5	Luminous Perceptibility	58
3.4	Spatial Cognition	60
3.4.1	2D vs 3D Interaction	61
3.4.2	Memory - The Use of Recall as a Measure of Visual Perception and Search	62
3.5	Discussion	64
4	Augmented Reality System Development	68
4.1	Introduction	68
4.2	Hardware	68
4.3	Tracking	69
4.4	Software - Windows Presentation Foundation (WPF)	70
4.5	Initial System Design Concept and Development	71
4.5.1	Edge Detection	72
4.5.1.1	Testing and Results	72
4.5.2	Automatic Machine Recognition	73
4.5.3	Neural Networks	75
4.5.3.1	BPN Applied to Pattern Recognition	76
4.6	Neural Network AR System Analysis	77
4.7	WPF Software Design	80

4.7.1	2D or 3D	81
4.8	3D Graphics Theory	82
4.8.1	3D Viewport	83
4.8.2	Direct Show and WPF	84
4.9	AR System Composition	85
4.10	Discussion	87
5	AR Performance Study - Frame of Reference	89
5.1	Introduction	89
5.2	Participants	91
5.3	Hardware Design	91
5.4	Software	92
5.5	Stimulus Objects	92
5.6	Experiment 1A	94
5.6.1	Design	94
5.6.2	Hypotheses	96
5.6.3	Procedure	97
5.6.4	Results	101
5.6.4.1	Measurements	101
5.6.4.2	Analysis	102
5.6.5	Discussion	108
5.7	Experiment 1B - Primacy and Recency Memory Strength: Blocking	110
5.7.1	Design	110
5.7.2	Hypotheses	112
5.7.3	Procedure	112
5.7.4	Results	114
5.7.4.1	Measurements	114
5.7.4.2	Analysis	115
5.7.5	Discussion	117
6	AR Performance Study - Immersive Reality vs Real Life	119
6.1	Introduction	119
6.2	Participants	120
6.3	Hardware Design and Software Development	120
6.4	Stimulus Objects	124
6.5	Design	125
6.6	Hypotheses	127
6.7	Procedure	127
6.8	Results	129
6.8.1	Measurements	129
6.8.2	Analysis	130
6.9	Discussion	135
7	AR Performance Study - Search of a Local Environment	138
7.1	Introduction	138
7.2	Participants	139
7.3	Hardware Design and Software Development	140

7.4	Stimulus Objects	141
7.5	Design	141
7.6	Hypotheses	144
7.7	Procedure	144
7.8	Results	147
	7.8.1 Measurements	147
	7.8.2 Analysis	148
7.9	Discussion	151
8	Conclusions	155
	8.1 Satisfaction of Problem Statement	155
	8.2 Further Work	159
	8.3 Final Thoughts	162
A	Methodology Diagram for Forensic Investigation Improved Through the Advent of Ubiquitous Computing Methods	164
B	Ultrasonic Transmitters Positioned Around a Real World Environ- ment	167
C	Neural Network Node Configuration with Relative Weights for Ziffer 1	169
D	Neural Network Based AR System Work Flow Diagram	171
E	User Consent Form	173
F	Recall Sheet - Experiment 1A	175
G	Recall Sheet - Experiment 1B, Blocking Effect	177
H	Recall Sheet - Experiment 2	179
I	Experiment 2 - Scan Path Analysis Data	181
J	Recall Sheet - Experiment 3	184
	Bibliography	186

List of Figures

1.1	The Ubiquitous Camera	10
1.2	The Ubiquitous Portable Digital Assistant	11
1.3	Concept Drawings for the Ubiquitous AR System	12
2.1	The RV Continuum	22
2.2	Invisible Trains, AR on a PDA	30
2.3	HMD Conceptual Diagram	32
2.4	Visualisation Plane, 6 DOF	34
2.5	Accelerometer based on MEMS technology	39
2.6	Typical Mechanical Structure of a Gyroscope	39
2.7	Gyroscope, measurement of resonating mass	40
2.8	Gyroscope, movement of resonating mass	40
2.9	A Typical Fiducial Marker	42
3.1	Visual Field of View, Monocular HMD	53
3.2	Visual Field of View, Binocular HMD	53
3.3	Human Eyes, Monoscopic and Stereoscopic Overlap	54
3.4	Human Eyes, Binocular Rivalry	55
3.5	Contrast Sensitivity Chart	59
4.1	Comparison of Edge Detection Operators	73
4.2	Neural Network Node and Layer Structure	76
4.3	Screenshots of the Pattern Recognition System	77
4.4	Sample output of NNAR to user view frustum	78
4.5	Pilot User Trial Scene	79
4.6	A Camera's Relative FOV	83
4.7	Cube Geometry in XAML	84
4.8	AR Software Architecture	86
5.1	Daeyang i-Visor	91
5.2	Creative Labs - Live! Ultra Webcam	92
5.3	Sample of 3D objects used in AR experiment	93
5.4	Virtual Stimuli in AR On compared with AR Off	96
5.5	User Trial Scene Shot	97
5.6	Movement Condition Experimental Design	99
5.7	Static Condition Experimental Design	101
5.8	Graph, AR On vs AR Off Movement Condition	105
5.9	Graph, AR On vs AR Off Static Condition	106
5.10	Graph, AR On, Motion ON vs Motion OFF	107

5.11	Graph, AR Off, Motion ON and Motion OFF	107
5.12	Blocking Condition - Three Objects Per Marker	111
5.13	Experiment 1 Snapshot, Blocking Condition	112
5.14	Movement Condition with Blocking Experimental Design	114
5.15	Graph, The Additive Effect to Recall using Blocking AR On	115
5.16	Graph, The Additive Effect to Recall using Blocking AR Off	116
6.1	HMD Hardware, LiteEye LE-700A	121
6.2	Synthetic Renditions of Real World Objects	122
6.3	Real World Workspace	122
6.4	VR Workspace	123
6.5	Real World Task Workspace	126
6.6	A Grid to Show Distribution of Bottles	127
6.7	Experimental Design	129
6.8	Graph to Show Dwell Times	130
6.9	Real Targets: Fixation time versus Probability of Recall	132
6.10	Augmented Reality Targets: Fixation time versus Probability of Recall	133
6.11	Virtual Reality Targets: Fixation time versus Probability of Recall	133
6.12	Scan Path Diagrams for RL, AR and VR	134
7.1	The X^4 (Chi-Four) Wearable Computer	140
7.2	CSI Search Methods	142
7.3	Marked Synthetic Bottles	143
7.4	Synthetic Bottle in Scene	143
7.5	Plan of Floor Layout	145
7.6	User Trial Snapshot	146
7.7	Experimental Design	147
7.8	Graphs to Show Search Time and Recall Performance	148
7.9	Head Tracking Data, Yaw Tracking	149
7.10	Head Tracking Data, Pitch Tracking	150
7.11	Recall Performance against Object Position	151
A.1	Forensic Methodology Flow Diagram	165
B.1	Ultrasonic Transmitter Configuration	168
C.1	Ziffer Weights	170
D.1	Neural Network Based AR System Work Flow Diagram	172
E.1	Consent Form	174
F.1	Recall Sheet 1A	176
G.1	Recall Sheet 1B	178
H.1	Recall Sheet 2	180
I.1	Scan Path Analysis Data, AR Nodes	182
I.2	Scan Path Analysis Data, VR Nodes	182

I.3	Scan Path Analysis Data, RL Nodes	183
J.1	Recall Sheet 3	185

List of Tables

2.1	A Table to Show the performance characteristics of Low-High end HMD devices	33
5.1	A Table to Show the Synthetic Objects Used in User Trials	94
5.2	A Table to Show Average Recall	102
5.3	Experiment 1 - 2x2x12 Repeated Measures ANOVA	103
5.4	Experiment 1 - Test for Learning Effect	108
5.5	A Table to Show the Average Recall Using Blocking	116
5.6	Experiment 1B Blocking - 2x4 ANOVA	117
6.1	Experiment 2 Dwell Time - 3x12 ANOVA	130
6.2	Experiment 2 Recall Performance - 3x12 ANOVA	131
6.3	A Table to Show Average Recall	132

Dedication

In memory of

Stan Wood (Grampa)

Chapter 1

Introduction

1.1 Overview

As electronics hardware and algorithmic programming improve, emerging technologies such as Augmented Reality (AR) become more feasible to aid real world task solving. It should be noted that many studies of AR do not examine the ergonomic implications of systems, caused by factors such as device structure, weight, software lag and composite resolution restrictions within the user view. It is however, essential to understand how these properties inherent in AR systems affect human performance, as this will provide invaluable insight into how technology development, methods of use and contexts of implementation can best support human interaction with AR technology.

The research work in this thesis suggests that a system based on fiducial markers would form a basis for the most suitable system to evaluate ARs effects on human performance (see chapter 2, also chapter 4). Thus, the thesis is not concerned with the development of AR per se but rather with building prototypes which are sufficiently capable of replicating current AR performance in order to study how this affects human performance. In order to evaluate AR performance, augmented, virtual and real life reality scenarios are evaluated in comparable conditions. The first is the AR condition which comprises of an environment which is real and evidence (3D virtual objects) within it are synthetic. AR performance is then compared with equivalent real life scenarios (all evidence viewed through the head mounted display (HMD) is real) and Virtual Reality (VR) scenarios (environment and evidence viewed through the HMD is synthetic). Previous work (see

Chapter 3, 3.2.1) suggests that human performance is most affected by the translational component of movement in a real environment. As a consequence movement is the key motivation in proving ARs viability as a concept and as such is key to satisfying the research goals of this work.

Although a marker based system (refer to Chapter 2, 2.5 for a description) is deemed most appropriate for the research carried out herein, the advantages other tracking approaches could provide to a system for task based investigative work are not ignored. This system is evaluated and discussed in chapter 4 is comprised of a markerless approach built on the frame of a back propagation neural network. The neural network is first trained and then used to recognise real world objects of interest in real world space and augment additional contextual information. The trade-off on this system is computational expense, exhaustive neural network training and accuracy.

1.2 Problem Statement

The research in this thesis is concerned with the following two questions.

Can an emerging technology such as AR be utilised to aid problem solving methodology such as crime scene investigation?

What affects do the properties of AR have on human performance?

The scope of the thesis satisfies the research goals over three stages:

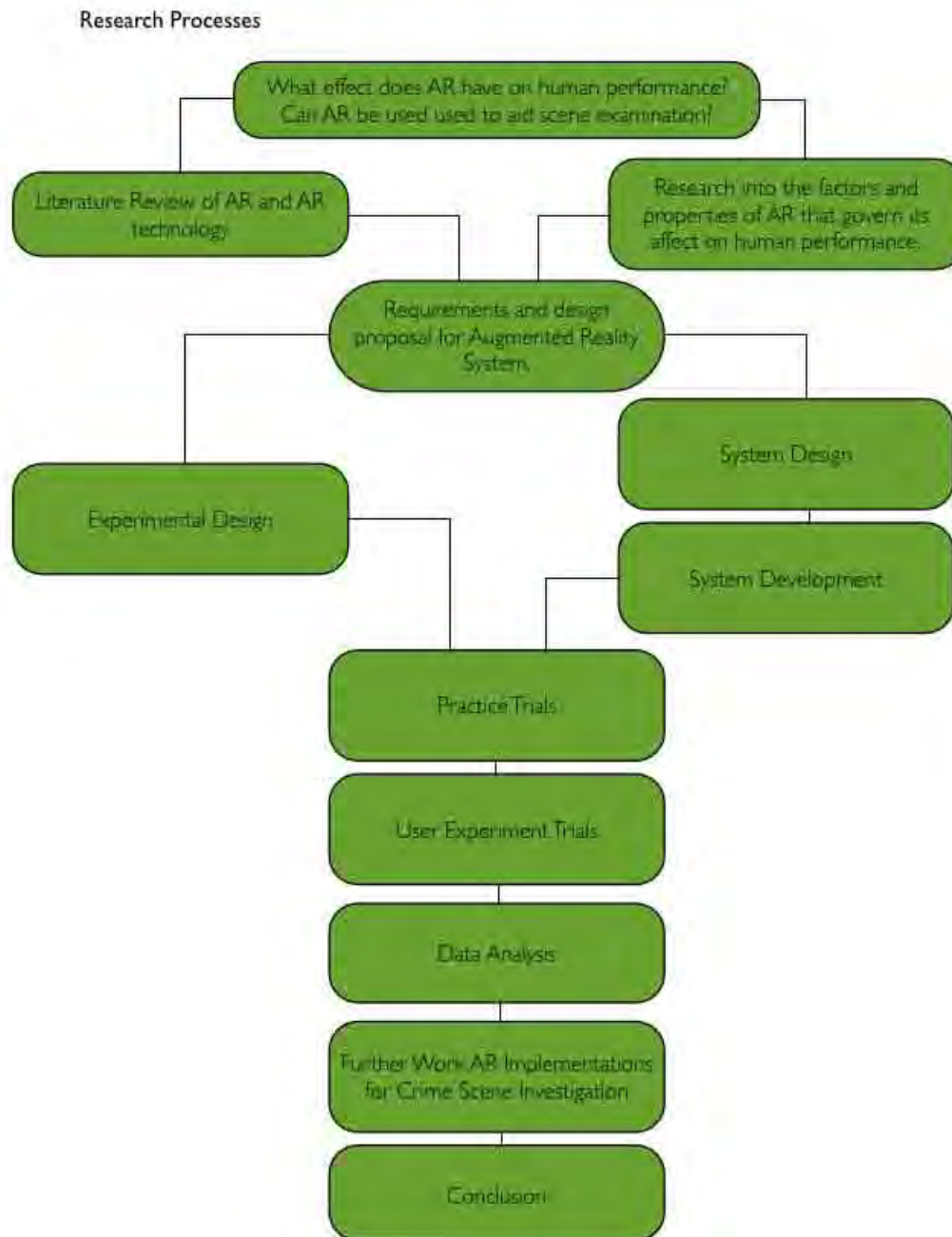
1. Research of available literature on technology and human performance related issues pertaining to the properties inherent in AR systems.
2. Design and utility of an AR system for human task based performance analysis.
3. Implementation of a series of experiments designed to evaluate the cognitive implications pertaining to the utility of synthetically enhanced environments for task based performance analysis in crime scene investigation.

1.3 Research Assumptions

The problem addressed in this thesis is concerned with the utility of AR for human task based work and the effect AR technology has on human performance. Existing literature on AR is primarily concerned with factors that govern the composition of AR technology such as tracking display methods, software design and novel contexts for application. This research looks at existing technology implementations and reviews a body of human perception factors that relate to the properties of AR technology.

1.4 Research Processes

Ideally the system developed for task based human performance research would be catered and implemented in a mock crime scene using forensic professionals as participants. This however, proved difficult to orchestrate in terms of logistics. It was therefore concluded that performance issues relating to AR would be researched using students at The University of Birmingham. Tasks were therefore designed such to be suitable for the participants involved and the forensic spin so to speak was removed. This however was believed to have no bearing on the research findings. Each task based trail was focused on the review and recall of individual items and the effects pertaining to AR properties had on performance.



1.5 Domain of Investigation

1.5.1 Synthetic Environments - Transfer of Training

The applications for augmented reality technology are vast ranging from education, gaming and entertainment to the medical and construction field. Most research suggests that the security and defense sector, particularly the Office of Naval Research and Defense Advanced Research Projects Agency or DARPA [1], are some of the original pioneers

of augmented reality systems. One of the main uses of augmented reality systems to the military is providing field soldiers crucial information about their surroundings as well as friendly troops and enemy movements in their particular area [2]. Another very important application for augmented reality systems is its potential role in law enforcement and intelligence agencies. A well-designed AR system could for example be used to aid law enforcement agents in a real world crime scene by synthetically re-implanting collected evidence into a scene that had been reopened for investigation (such as a cold case murder) helping the CSIs methodological investigative and narrative process (as explored in experiment 3, chapter 7). Another useful application for AR technology could be to append evidence with additional context sensitive information arrays (see chapter 4).

1.5.2 Crime Scene Investigation

Broadly speaking, crime scene investigation begins with an incident that can be interpreted as criminal, proceeds through to examination of a scene, to the selection, collection and analysis of evidence, and to relating the evidence to a case that can be answered [3]. The crime scene itself therefore constitutes a location and a plot. The crime scene is a location that has existed in a particular state at a certain point in time and this place is the habitual landmark for the scene of some kind of criminal activity. As such, this location will have been encoded with a variety of marks and traces that can be interpreted and discovered. Traces of blood, nails and hair constitute (DNA) codes, which can be decrypted and deciphered and in the same way gunshot residue (GSR), bullet holes or physical damage are signs that can be read and interpreted. The remediation of a scene of crime therefore, is a case of building up a narrative that concludes with a criminal action. At first this narrative is hidden and scattered and so has to be revealed via a methodological process of investigation. The forensic detective's ability therefore to piece the narrative of a crime scene together using various forensic methods, logical reasoning and deductive thinking is crucial to successfully solving a crime [4].

1.5.3 Case Solving Methodology

The purpose of examining a crime scene is to formulate a hypothesis based on all the available evidence understanding the most likely course of events that has resulted in

the observed circumstance [5]. Within the domain of crime scene investigation there are four main types of resource for action that are used to construct a narrative for a case.

- The environment - this affords differing forms of examination.
- Objects within the environment - these require a certain interpretation as evidence.
- Procedures that govern the investigative process - this affords the type of application in different environments.
- Narrative constructed during the course of investigation - this consists of hypothesis formation and explanatory models of the incident.

Using these resources a crime scene investigator builds a case that can be tied in a court of law. The crime scene investigators role in case solving methodology is therefore that of search and retrieval of evidence in a scene of crime. How much the activity of a crime scene investigator can be considered as search and how much it is a retrieval activity is a subject of debate within the CSI community. One school of thought argues search and therefore interpretation of evidence in a scene of crime is essential to their role in determining a narrative of the crime committed. This leads to the argument of how much leeway should be granted to a CSI in terms of interpreting an object's evidential value in a case. As such ignoring the obvious consideration of how retrieval can even be performed in the absence of search one school of thought suggests a CSI's role is to collect and recover evidence in a manner that is as neutral as possible and leave interpretation to other specialists [6]. What is interesting is that the range of systems currently employed by CSI tends to support the latter school of thought.

In the UK there are currently three commercially available systems to aid crime scene investigation. These evidence management systems allow the transfer of evidence from crime scene to laboratory to courthouse to be performed reliably. The SETS1 (Single Evidential Tracking System) can be used at the crime scene and supports recording of scene of crime details, Modus Operandi, offences, found exhibits, and Forensic Science Service submissions. Anites Socrates2 system is a suite of evidence tracking and management systems that not only record information from the crime scene and tracks evidence, but also manages workflow and submissions. Locard3 uses a barcode reader (interfaced with a laptop computer) to read in the bar code printed on a particular item

of evidence that can be linked to its unique ID. All the commercial systems have been designed to link with some of the police computer systems, such as Holmes2 [7].

It is interesting to note how these systems have approached the problem from slightly different angles. Whilst they support the digital representation and tracking of evidence, the manner in which a digital identifier is assigned to an item of evidence differs between systems, e.g., Locard3 directly pairs the evidence bag with its digital identifier through the use of barcodes. Furthermore, the manner in which the system supports the overall activity of managing crime scene investigation differs, e.g., Socrates2 provides support for managing the workflow of several different forms of investigation, e.g., scene of crime, fingerprint etc. Near-future developments of these systems and others appear to be directed at shortening the time taken between material being collected and analysed, and a suspect identified. To this end, there are several projects that use digital imaging to capture finger-marks or footway marks, and then use these images for analysis. The advantage of such approaches is that the material is digitised and can be sent wirelessly to the analyst.

These current and near future developments are no doubt helpful in advancing the turnaround and verification procedures of evidence collection. However, they also bring with them a controversial argument about the risks of over reliance on technology because they focus on a limited aspect of crime scene investigation, namely evidence collection. Such is well documented by a number of accounts in McCartney [8] and Findlay and Grix [9] regarding the impact DNA evidence has had on crime solving methodology.

You can slip into the lazy approach that we've got the DNA we needn't bother doing the rest of the work. What it does though is give you a concrete line of inquiry, which still need corroborating with other evidence. There is a lot of good old-fashioned detective work also needed [8].

This erosion of rights is perhaps most clearly evidenced in the recent cases of mass testing in Wee Waa and Norfolk Island, where non-compliance became not so much an exercise of choice but rather an act equated with an inference of guilt ... Arguably, this is best characterised by the familiar question heralded in media reports of the time 'Why wouldn't he give a sample if he had nothing to hide?' There has been enough challenge to the reality of informed consent within forensic procedures without the added strain concerned with the actuality of volition in mass-testing situations [9].

There are problems then with the training of officers in forensic awareness and dangers of a silver bullet mentality. Increasingly faith is placed in forensic science alone to fulfil a supporting or verification role in crime scene investigations, however, forensic science may serve to hide from the critical gaze detection where forensic evidence has been afforded apparent credibility, leaving the process of detection, evidence gathering and investigation hidden. The canopy of science obscures the primitive analytic tools that persist.

In the work of Schraagen and Leijenhorst [10] it is suggested that forensic investigators rely on a narrative to determine how best to collect evidence. This narrative might include expectations concerning evidence and expectations concerning the crime. In this study (in which trainee forensic investigators work at a simulated crime scene), they showed that experts use this narrative to select appropriate heuristics to guide their search for, and interpretation of evidence. The process by which items are selected can be considered analogous to directed search. This implies that search involves not only seeing evidence but also developing an expectation that something will be present. This inferred knowledge as a manner of influencing the course of search within a scene comes through experience in the field. Thus a system that enables distribution of knowledge about a crime scene could be invaluable in assisting the transfer of socio-contextual and spatial inference between expert level 4-5 forensic investigators to trainee level one officers thus improving their search routines and methods. Such could be achieved using AR. It could be possible to augment a scene with virtual representations of evidence enhanced with additional relevant information pertaining to an active case in an actual real scene of crime. Having evidence virtually represented and enhanced with relevant details in situ at a crime scene could then be used to build a narrative, lead investigators to new evidence or information and help investigators corroborate information.

1.5.4 Early Work - Proposed Augmented Reality Ubiquitous Computing System Concept for Evidence Reconstruction at a Scene of Crime

Following talks with an ex-forensic science services employee, insight was gained as to the methodology of a crime scene investigator. Using this knowledge, usability testing requirements have been hypothesized. Appendix A speculates how an Ubiquitous

Computing ¹ system that utilises AR could aid the forensic investigative process. The usability requirements for a ubiquitous computing AR system aim to automate and enhance the evidence documenting process, provide a digital medium for accessing crime scene evidence and using AR create an interactive visual interface to gain improved contextual awareness and better understanding of evidence on re-visits to the scene of crime. There are therefore two stages required for the successful implementation of this system, evidence retrieval and evidence representation.

1.5.4.1 Evidence Retrieval

The fundamental key to creating technology for human interaction is to make it interact with humans seamlessly. This is the motivation behind the ubiquitous camera, the hope is that it will be easily adopted by crime scene investigators because it will not interfere with their current methodology for evidence collection (see figure 1.1).

Orientation tracking technology (see chapter 2 section 2.4.5) and a distance/range calculator (see chapter 2, section 2.4.1) is affixed to the camera device to calculate the angular orientation and the distance from the evidence being photographed, co-ordinate information is recorded also using ultrasonic sensors (see chapter 2, section 2.4.3) placed in the room before evidence collection takes place. Using sound waves to determine distances is a cheap and effective method. Ultrasonic transmitters are placed in fixed positions around a room (see appendix B), the device to be tracked is fitted with an ultrasonic receiver that can measure the distance to each of the sensors. Knowing the distance to each sensor at any point in space the tracked objects position can be calculated very accurately [12]. Finally a portable computer with wireless technology is used to store photos appended with all 6 degrees of positional information. Back at the crime lab photos and data are loaded into the system which renders 3D models of each piece of evidence and allows forensic analysts to add relevant information to it.

¹Ubiquitous computing is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activities. In the course of ordinary activities, someone 'using' ubiquitous computing engages many computational devices and systems simultaneously, and may not necessarily even be aware that they are doing so. This model is described in Weiser [11] as the third and final paradigm as an advancement from the current desktop paradigm.

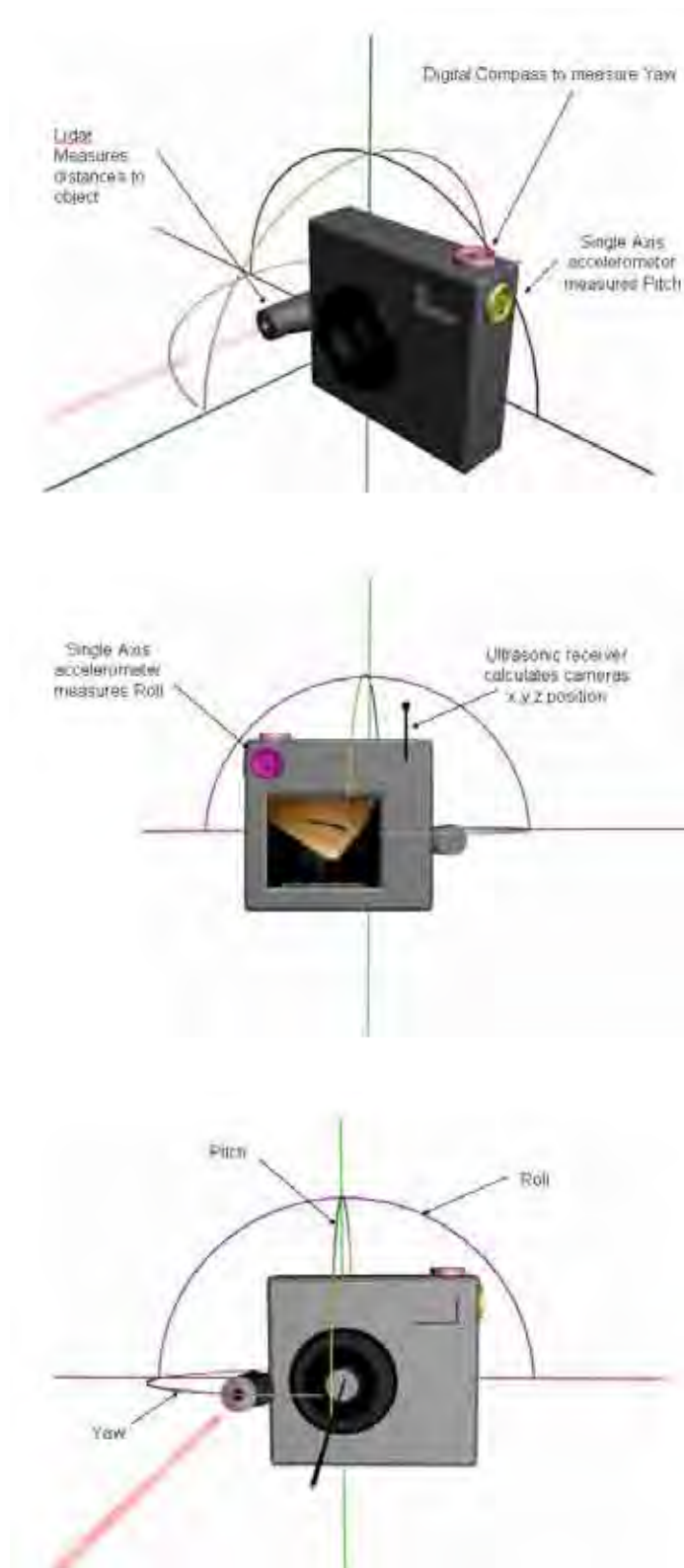


FIGURE 1.1: Ubiquitous Camera Concept Visualisation - A camera real world pose is estimated using a Lidar, ultrasonic sensors (possibly GPS for outdoor) and accelerometers.

1.5.4.2 Evidence Representation



FIGURE 1.2: Ubiquitous User Interface

Reconstruction at the scene of crime is based primarily around a touch screen implementation (see figure 1.2) although it will be possible to adapt this system to be used with a head mounted display (see chapter 2, section 2.3.2). The screen shows the user the real scene augmented with virtual information (see figure 1.3). The real scene is captured via a camera device and chroma keying techniques are used to place virtual evidence in the video rendered real scene. Like the ubiquitous camera the co-ordinate position of the PDA is registered in real space using a mobile ultrasonic positioning system placed at predefined positions within the environment.

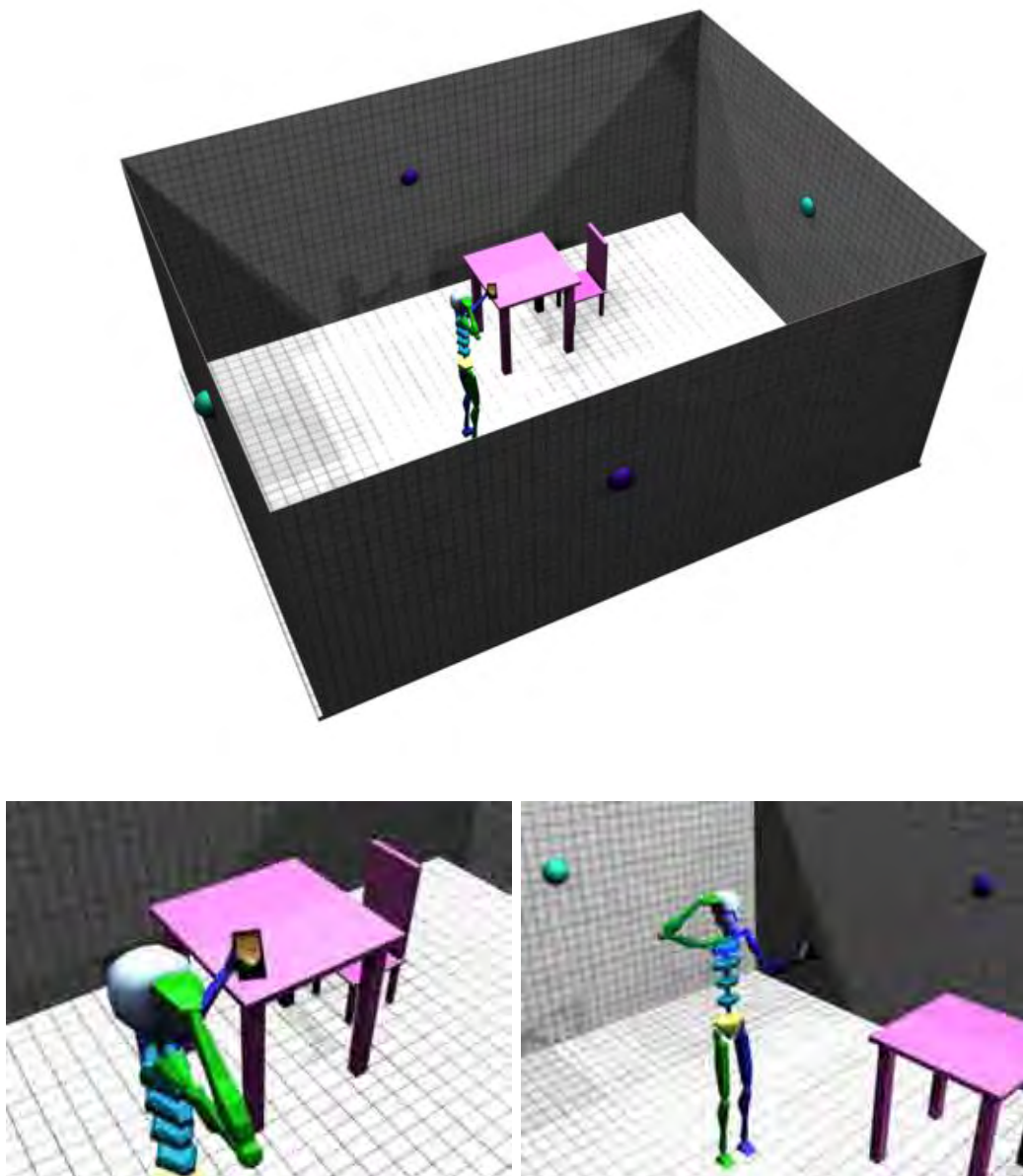


FIGURE 1.3: Concept, Virtual Evidence Represented in Crime Scene. The user sees a synthetically enhanced view of the world through the PDA. The knife that the user sees is an example of evidence removed for examination that, using AR can now be synthetically re-placed at it's original real world habitual landmark.

1.5.4.3 Concluding Thoughts

The hope and ambition motivating this project was to develop a completely immersive mixed reality system for the user, ie information is augmented in real time as the user moves around in the world. For such a system to be realised a myriad of technologies would need to be brought together to operate in unison. The system proposed revolves around the development of a ubiquitous camera (acquisition) and a ubiquitous visual interface (representation). There are substantial and interesting technical challenges arising from these ideas. However, it was felt that simply pressing ahead to address these challenges was not appropriate, given our current understanding of how people relate to AR. Thus, the focus of the research shifted away from a primary concern with technology and towards the scoping of basic human factors considerations and challenges. The development of a more mainstream AR system held more value in regards to its merit in proving AR as a concept in its more widely adoptable mainstream form. Generally speaking the use of visual fiducials (see chapter 2, section 2.5) is the most commonly used implementation of AR and therefore any studies relating to human performance would hold more merit within the research community. Although there is a paucity of research relating to human performance using AR, there is however research to suggest the utility of synthetic environments can be used to improve human performance.

1.5.5 Transfer of Training - Virtual Environments to the Real World

The previous section suggests that transfer from search in VR platforms to real world environments depend heavily on the constraints governing the task. In VR representations of flight simulations, a successful paradigm between transfer from VR training to real world performance has been directly associated. Flight simulators seek to support the same actions as are found in a real environment and allow these actions to be performed under similar conditions, such is achieved over a wide variety of media from applications that run on home computers to entire rooms built to be exact replicas of real planes. In these replicas, all controls and electronics are made exactly as they would be found in the real plane and behind the windshield are high resolution displays made to approximate what the pilot would see in a real plane. In some highly advanced systems turbulence and G-force is also simulated using hydraulics.

Despite the fact that the employment of motion may enrich realism in the VR experience, research into the transfer of training from flight simulators to actual flying has shown that the inclusion of motion in VR flight simulation adds little to performance in real world flying. All contact, formation, navigation, instrument, air-to-air and air-to-surface tasks can be taught effectively, quickly and efficiently using economical VR simulations. This is a consequence of the fundamentally cognitive nature of flying. Hence VR can be used effectively where the requirement of technology is closely aligned with the requirement of the user. Other cases where VR is proving to be effective is in driving simulations and training for surgical procedures. Seymour et al. [13] showed that gallbladder dissection was 29% faster for VR trained residents, it was also concluded that non-VR trained residents were 5 times more likely to injure the gallbladder or burn non-target tissue and 9 times more likely to transiently fail to make any progress. Moving over to driving simulators de Winter et al. [14] conclude that there was a higher chance of passing the driving test first time if the user made less steering errors in the VR simulator. Conversely earlier research by Roenker et al. [15] concluded that simple speed-of-processing tests provided benefits to driving performance that a VR simulation did not suggesting that VR is by no means a panoptic substitute for conventional methods. Although the findings in this body of research shows promise for simulated training neither ignore that there is still a paucity of research pertaining to these areas. Both Seymour et al and Winter et al conclude that further research is necessary on a larger scale to inarguably prove VR as a viable concept to provide transfer of training to the real world.

There have been various studies conducted that evaluate the utility of crime scenes reconstructed as virtual environments. The methods adopted focus on two aspects; speed and realism. Gibson et al. [16] propose a system that constructs complex and accurate three-dimensional models of real world environments from video sequences captured with standard consumer-level digital cameras. See also, an earlier example using photographs [17]. In another study Murta et al. [18], a myriad of technologies are used describe how accurate realistic virtual scenes can be created. This is achieved using a scene description language to construct complex geometrical models; the application of high-quality radiosity algorithms to render light paths and texture extraction from forensic photography. The merit of this kind of research can be recognized in the industry. Crime-houses worldwide are adopting VR packages to circumvent traditional methods

of imparting knowledge pertaining to forensic investigation.

VR packages are used not only to train but also to test users, these packages explain how to collect and preserve evidence and simulate crime scenes related to burglaries, explosions, homicides, firearm usage, DNA related cases, narcotic drugs, food adulteration, document fraud and computer forensics. Commercially available packages being adopted by police forces worldwide include CrimesceneNET² and CSV³ and in the U.K., a collaboration with the Greater Manchester Police can be seen in Howard et al. [19]. Unfortunately no commercially available package provides the user with an immersive VR experience of the scene, simple point and click is still the norm. Referring back to section 1.5.3 showed that crime scene investigators walk around a crime scene in order to build up a narrative. Thus, there is a need for simulated training environments to facilitate physical movement. So while these aforementioned packages that facilitate search exclusively may assist in training on the rudimentary methodology of collecting evidence and scanning a scene, they also serve to present distractors when transferring to real or mock scene of crime investigation by inadvertently skewing search cues that would otherwise be facilitated by movement. Darken and Banker [20] look at transfer of strategies from VR to real worlds and concludes that strategies and techniques that resulted in a reliance on perceptual imagery rather than symbolic information tend to cause problems.

Even with a VE that is relatively rich in information useful for navigation, it is still not close enough to the real world for a user to resolve the differences. Consequently, participants tend to believe they are not in the correct place in the real world because it does not match what they saw in the VE. These behaviors were not present in the map group participants as they encoded symbolic information only and therefore were not confused by detail. [20]

This begs the question of how best to combine mobility and search using synthetic media. Since movement is an inherent factor in AR systems its utility to perform search tasks could potentially circumvent issues of transfer that are problematic of VR systems that simulate movement.

²<http://www.crimescenenet.com/>

³<http://www.crime-scene-vr.com/>

1.5.6 Discussion

The salient deduction to be made here is that simulated environments transfer well in the real world when the requirement for the technology is closely aligned with the requirement of the user. Suffice to say in all successful implementations of VR training environments the real world task is limited to a constrained environment. In essence flying an aeroplane in the real world is a restricted environment consisting of linear controls that alter the the planes behavior across a two dimensional view frustum, similarly driving a car abstracts complex real world movement into a set of even less linear controls, forwards or backwards (via pedals) and left and right (via a steering wheel). In addition surgical simulation while fundamentally more complex to render synthetically (because of the requirement to register complex hand-arm gestures) the environment is no less constrained, specialised and detached from natural behavior in the real world. What these VR simulators do prove is that if technology can be replicated and designed to behave in the same way in a virtual world as in the real world then VR can be used successfully to train human beings to control that technology. What current research does not deduce is how human behavior is actually affected by the constraints this technology imposes on them. The glut of work in the domain of mixed-reality (MR) is focused on providing a user with a rich experience and evaluating that experience based on subjective measures of performance such as the degree of human interaction or the perceived mimicking of human behaviour. Unless we can measure the fundamental aspects of human behavior in MR environments it is not possible to deduce whether that experience is a true experience or not. Hence further research and an involved set of psychometric testing are needed to deduce whether or not human behaviour and thus performance is affected by the constraints typical MR hardware imposes on the user. Indeed if there is conclusive evidence to suggest human behaviour in AR environments is not comparable to a real equivalent then one can only deduce that AR is not suitable as a tool to evaluate or assist human performance. Thus, the use of todays AR technology as a training model or investigative tool for crime scene investigation would be deemed inappropriate as an alternative or addition to conventional methods of training and investigation.

The core role of a CSI at a scene of crime is to search the scene, examine evidence and build a narrative for the crime. From the results of this crime assessment a CSI will

build up a profile of the most likely assailant that will hopefully lead police detectives to the suspected perpetrator. Using VR environments it could be possible to reconstruct a crime scene and facilitate navigation. However as discussed earlier, research indicates that transfer from VEs that support body based movement to real environments is poor and this is thought to be because the user requirement is not well supported by the technology. Where the requirement of the user and the technology is closely aligned transfer has proved to be good. Thus, in a well conceived implementation of VR or AR media where the requirement of the user is satisfied, a highly effective training domain platform can be expected. Although the implementation of an AR system such as the system proposed in this chapter is fairly palpable the research in this chapter suggests that the alignment between real life and synthetic media needs to be well analysed before a multi modal technology such as AR can be deemed appropriate for human training. Overall two key aspects pertaining to VR simulation seem prevalent.

- Physical body movement by the user around a virtual environment is inherently lacking in most implementations of VR technology and consequently when this key principle is facilitated transfer of training to real life is compromised.
- Training using VR works well when the requirement of the technology is closely aligned with the requirement of the user, such as aeronautical and driving simulators.

These two most pertinent findings in the domain of VR research suggest that when user interaction is satisfactorily comparable between conditions good user transfer can be observed. Where such success is observed can be described as the technology providing the user with a true-to-life experience or simply 'true experience'. A principle factor limiting this in most guises of VR is the ability to facilitate movement. Thus it would seem reasonable to deduce at this stage that a series of tasks that analyse varying degrees of movement, would be prudent in identifying how successful AR can be expected to be at providing a true experience.

Investigation at a scene of crime can be measured primarily as a search task. If behavior when performing search in various modes of reality such as VR and AR differs completely to real world behavior, then the prevalent deduction to be made is that the technology does not support normal human behavior. A further remark then, would be that these

technologies are not suitable for tasks other than the specialised expert control systems described earlier. In order to model human behavior and measure human performance efficiently and effectively factors pertaining to human perception of real and virtual environments should be carefully considered. This is achieved in the following work through a series of experiments that begin by looking at how the perception of movement affects performance in a scenario where the requirement of the technology is aligned exactly with that of the user. This series of experiments works towards incrementally increasing the requirement of technology (by adding extra modals of movement) and as such, proves that as the task becomes more complex for the technology the suitability of AR to train and measure human performance diminishes.

1.6 Contributions Made

This thesis contributes to research within the field of Augmented Reality by:

- Developing an AR system that can be used to research human based task performance.
- Detailing the cause and effect that properties of AR pose to human performance.
- Assessing whether AR is a viable platform for implementations of novel approaches to human based tasks that could be used to aid spatial navigation and recall.
- Summarising that where the requirement of the user is closely aligned with the requirement of the technology AR is a very suitable equivalent to real life scenarios. However, as this close correspondence of user and technology requirements reduces so does the suitability of the AR platform to provide a true user experience.
- Showing that the navigational component of movement has a dramatic effect on human performance. Since movement is inherent and unique to AR technology, proves the value of its utility in task based search. In addition to this the findings in experiment 3 (see chapter 7) show that more advanced systems do need to be developed in order to facilitate full body based movement in AR.

1.7 Thesis Structure

Chapter 2 provides the reader with a background to augmented reality and defines how AR fits into the field of mobile computing. It also provides a foundation for the reader to relate to and appreciate what AR technology is and how it relates to other computing technology such as Virtual Reality and Augmented Virtuality.

Chapter 3 reviews aspects of AR relating to its technological implementation and re-searches a key property of AR, that being to alter visual perception. In this chapter a review of literature on AR technology is provided that focuses on how the properties of this technology affect visual perception. This chapter attempts to facilitate an understanding of how human performance could be affected by the relative bias technology employed by an AR system poses to visual perception.

Chapter 4 details the development of an augmented reality system that is utilised to evaluate task based human performance. Justification of design choices are detailed and the concept and design process for a system focused on evidence review and analysis is explained. A system is designed that can be trained to recognise real world 3D objects. This is then used to augment the real object with additional information. However due to the computational complexity that such a system would impose to registration accuracy in a dynamic environment the system is developed such to utilise a fiducial marker based system instead. The system developed is based on the popular and well documented ARToolkit which is integrated into the .NET framework allowing it to benefit from the graphics capabilities of the Windows Presentation Foundation namely it's graphics core, called Windows Graphics Foundation (WGF).

Chapter 5 is the first experiment chapter. This details the practical design, procedure and analysis of data from an experiment that looks at the effect that AR properties have on human performance. The experiment details 2 conditions that comprise movement ON/OFF with AR ON/OFF. In the movement condition 2 states are observed; one where synthetic objects are reviewed without a background (called AR Off) and a second where synthetic objects are presented with the real world view available (called AR On). In both states movement across only one (the rotational) axis plane is facilitated. The user rotates about 360° to review 12 objects positioned per clockface. The static condition investigates how human performance is affected when movement

is not facilitated and once again comprises of AR ON and AR OFF states. A further experiment looks at the effect that blocking of objects together in sets of three has on human memory recall performance.

Chapter 6 makes the user requirement slightly more advanced in that full rotational head based movement is facilitated. In this trial three conditions are facilitated, a real life condition, an augmented reality condition (synthetic objects placed over markers), and a virtual reality condition (fully synthetic environment). Movement in the VR condition is facilitated using head tracking. Head tracking data is recorded in all conditions and specific information is extracted from each trial such as dwell time and scan path.

Chapter 7 brings the utility of AR to a fully facilitated wearable platform allowing for a full six degrees of freedom across the translational and orientational co-ordinate reference frame. A scene is created in which users are asked to investigate and search for objects of interest. Two conditions are evaluated, real life where objects to be located are real and AR where synthetic equivalents of the objects are presented to the user. This chapter attempts to conclude ARs justification as a viable platform for human based task work and takes the findings from previous experiments further by analysing the effects of the translational component of movement.

Chapter 8 summarises the main findings and contributions of the research with regards to the original problem statement. Suggestions related to directions for the further work and evaluation methods pertaining to AR technology and its inherent design properties are also addressed.

Chapter 2

Reality and Virtuality - An Introduction to AR

2.1 Introduction

The research presented in this thesis is concerned with the use of technology to evaluate how its application affects a specific aspect of human performance, namely perception and memory. This is achieved through the development of a system that combines computer generated imagery with the real world by way of visual fiducials. All of the currently available media of computer generated 'reality' such as the combination of virtual information with the the real world can be defined under a class of technologies called Mixed Reality (MR). There are a number of factors to consider when categorising various MR systems. However, generally MR systems can be categorised as a) Augmented Reality or b) Augmented Virtuality [21].



FIGURE 2.1: A Simplified Representation of the RV Continuum. Reproduced from Milgram et al. [21]

Figure 2.1 that shows the reality-virtuality (RV) continuum. In the case of the left of the continuum we define an environment which consists solely of real objects, this includes anything that may be observed when viewing a real-world scene either directly in person, or through some kind of video display.

The use of a video display requires that the user's view of the world is mediated by a viewing device such as a camera. Vision is thus, limited to the optical power of the device and determined by factors such as display resolution, field of view and contrast ratio (see chapter 3). This means that even when viewing a real scene through technology there is likely to be some attenuation and selection in what is being seen. This is important to note because the presentation of virtual media, using display technologies mixed with a view of the real world, could increase these affects of attenuation and selection further. The case at the right of the continuum (see figure 2.1) defines environments consisting solely of virtual objects, which includes conventional computer graphic simulations, monitor-based or immersive. Conventionally, such media are termed virtual reality [22] and rely on the construction of entirely synthetic views of the world. The visual element can be achieved using sophisticated drawing packages, such as 3D Studio Max and 3D rendering engines such as the Crytek engine, (both of which are utilised in chapter 5 to create a virtual reality environment). Movement in the synthetic environment can be simulated by providing visual stimulus via a head mounted display and tracking user head movement using three or six degree of freedom tracking devices such as the Wiimote (used in experiment 2, see chapter 5), Intersense iTrax (used in experiment 3, see chapter 6) or Magnetic Trackers such as the Polhemus Patriot (see section 2.4.4).

The framework described here shows the relationship between a virtual environment and a real environment. Generally the goal when utilising any form of MR is to create a

symbiosis between real and virtual environments. At first blush then the task of creating AR imagery that can be accurately perceived may appear to be simply a matter of synthesising one or more perceptual cues, such as relative size, and merging it with a view of the real world. That it is possible to produce trompe l'oeil¹ implies that synthetic cues can be created, but at the same time the so-called argument from illusion also implies that these cues can badly mislead the eye with regard to ground truth. Richard Gregory refers to 'perceptions as hypothesis' for the following reason:

Perception is not determined simply by the available stimulus patterns; rather it is a dynamic searching for the best interpretation of the available data... The senses do not give us a picture of the world directly; rather they provide evidence for checking hypotheses about what lies before us. Indeed, we may say that a perceived object is a hypothesis... [23].

Therefore, difficulties may lie in the way in which the brain combines various cues to maintain consistency across a scene as a whole. Thus, the problem of realistic registration in MR environments is not easily overcome. In-fact with current technology it is not possible to provide indecipherable MR. If such were possible it may be fair to argue that human performance would be unaffected by these states of mixed reality since the whole scene would be perceived as real. We have already noted earlier that the very act of viewing the world through some kind of display technology, say a camera, could be sufficient to change the viewer's perception of the world. This delicate nature of human perception can be observed in reality where natural phenomena such as a heat mirage can deceive human perception. Heat mirage is caused when atmospheric refraction by a layer of hot air distorts or inverts reflections of distant objects creating an optical illusion. This means that, in effect, the brain's interpretation of the optical information from the eye can be at odds with what is actually there. This is of course the basis of visual illusion, and these basic processes of perception need to be considered in order to appreciate the points being raised here. This distinction between real and virtual worlds is important in determining how human performance is affected when utilising MR technologies. A detailed analysis of visual perception as it relates to AR and human performance is presented in chapter 3.

¹french for "trick the eye, is an art technique involving extremely realistic imagery in order to create the optical illusion that the depicted objects appear in three-dimensions, instead of actually being a two-dimensional painting.

2.2 RV - Combining Real and Synthetic Media

The reality - virtuality continuum comprises many forms of technology to achieve the designer intent. The following section considers this technology in their most commonly adopted applications.

2.2.1 Virtual Reality

Virtual reality quite simply describes an environment that is composed solely of virtually generated information. Virtually generated environments have become widely adopted mediums for human interaction and this technology has been used extensively in the commercial and research sector. Examples include internet based MORGs (massively online Role-playing game) such as World of Warcraft and Simple Life. Where these virtual environments fall short is the way in which the user interacts with them. Currently the mouse and keyboard or games pad are the most common user interfaces. Fully immersive VR is characterised by the interface and not just the content [24] and this is key to creating environments that support natural interaction². The ultimate goal in any immersive VR experience is to create a world which the user interacts with and perceives as if it were the real world. Immersion is also very much a response of the person [25]; one could be immersed in a book, a film or a simple video game. This means that immersion could be seen as a matter of managing attention to the VR and reducing distraction, for example; irrelevant features in a virtual environment [26].

Ideally then immersive virtual reality would be achieved through an exo-centric approach. Sadly replicators, force-fields, holograms and transporters are not available to us. Current VR has to rely on technology that monitors humans endo-centrally using trackers to register movement and head mounted displays to render the environment.

The visual effectiveness of a VR environment depends heavily on the 3D engine employed or created by the developer. Developers of VR systems rely on improvements made in the commercial sector to offer improved and more realistic graphical environments. The problem this presents to VR developers is that game engines such as Half Life (Source engine) or Far Cry (Crytek engine) generally do not support real world tracking devices

²Natural Interaction is defined in terms of experience; people naturally communicate through expressions, physical gestures and discover the world by looking around and manipulating physical objects.

natively and are closed source. This means custom methods of interfacing tracking hardware have to be developed. Rather than inserting three dimensional tracking co-ordinates directly into the virtual world, for example often tracking pose will be translated to emulate keyboard and mouse inputs. This method of development is generally buggy and cumbersome in its design using workarounds and non conventional program dll injection routines. There are however, many software development kits (SDKs) available that allow for bespoke VR environments to be created. Examples include Orealia, Quest3D and Optitrack to name a few. Costing of these platforms is reasonable and license holders usually get the benefits of a general public license (GPL) so code can be modified, implemented and distributed without the risk of punitive action. The utilisation of VR SDKs is the favoured approach for most VR system development but often at the cost of realism. This is because commercial games engines offer cutting edge physics and high resolution textured 3D environments with realistic shadowing, lighting and elemental effects that cannot be easily replicated in affordable development SDKs.

2.2.2 Augmented Reality

Augmented reality (AR) merges 3D virtual objects into a real environment and then in real time displays this augmented image to the user. Unlike virtual reality which seeks to immerse the user in a synthetic environment, AR supplements the real world with synthetic information. This makes AR the perfect tool to aid and even enhance human perception and interaction. Areas where AR technology is being applied widely varies from first person AR indoor and outdoor gaming engines [27], AR supplementals for vision based learning and tracking and augmentation overlays of patient information to aid surgery [28] [29].

It should be noted that thus far no normative terms of definition have been arrived at and since augmented reality is a relatively new technology there are still issues and debates surrounding the terminology relating to it. An example where the definition of AR is subject to debate is its use to enhance the viewer experience in live television events, such as realtime virtual corporate logo product placement on cricket and rugby fields. One could argue this is AR because virtual information is being rendered in real time onto a real environment. The reason why this definition of AR is contestable is in the way the user interacts with the mixed reality scenario. A standard visual medium such

as a television provides no body-based interactive method of interfacing and therefore by all normative terminology, is not actually augmented reality. The following denotes what is probably the most widely accepted and best illustrated definition of augmented reality:

In contrast with virtual reality, which refers to a situation in which the goal is to immerse a user in a completely synthetic environment, augmented reality refers to a situation in which the goal is to supplement a users perception of the real world through the addition of virtual objects. Azuma [30]

Azuma [30] also declares that a system can be described as being Augmented Reality by the following criteria:

- combines real and virtual elements.
- is interactive in real-time.
- is registered in three dimensions.

As such immersive augmented reality is an emerging genre of technology which aims to enhance a users perception of and interaction with the real world. Typically this is achieved through the use of a head mounted display which provides a medium to augment virtual data to. In order to augment virtual data into a user's viewpoint such that registration with the real world is accurate a user's position in the real world requires tracking. The most popular approaches include:

- The use of a head-mounted camera coupled with image recognition software to detect markers in the real world. This could use either a visual display to superimpose virtual objects on a real scene or to tie the virtual objects to items in the frames in a video feed.
- Orientation and translational registration methods that use tracking technology in various guises to monitor a users full range of body and head motion over a six degree plane in the real world. In many cases tracking methods are combined to achieve accurate registration; the ARQuake project [31] for example implements a combination of GPS, inclinometers, digital compasses and pattern recognition to achieve registration. Accurate registration however, is no easy feat to overcome

and a number of issues regarding the accuracy of registration in this system are highlighted in Thomas et al. [32].

The work of Milgram et al. [21] reviews a taxonomy of head mounted display units. In both dimensions of the proposed taxonomy some general issues within the context of MR environments are discussed in which multisensory stimuli have a main role. The following depicts AR scenarios in which other modalities are involved:

- **Auditory AR:** environments in which sounds from the real world and synthetic spatialised (virtual) sounds are mixed together. Research carried out by Wellner et al. [33] employs such a system.
- **Haptic AR:** environments in which information related to touch and pressure is superimposed on existing haptic sensations: for example, virtual objects can be touched by employing special kinds of glove devices. One manifestation of this is the use of a projection augmented (PA) model³ as is employed in the work by Bennett and Stevens [34].
- **Vestibular AR:** environments in which information about the acceleration of the participant's body in a virtual environment is superimposed to existing ambient gravitational forces (as, for example, in commercial and military flight simulators).

A theme that recurs throughout the literature on MR is the issue of stimulus combination; between real and synthetic imagery, between distance cues and between modalities (visual and auditory). This is perhaps not surprising given the *raison d'être* for AR/AV technology is ultimately to present synthetic stimuli that can be integrated with the real world. Stimulus combination can have various outcomes; it can fail, it can be used to enhance perception in a synergistic manner and it can be used to disambiguate ambiguous perceptual data so that it forms a robust percept. Given then that stimulus combination can be both a risk factor and an opportunity for the enhancement of task performance (either through improving efficiency and accuracy or reducing failure) this issue should necessarily form the backbone of a scientific research programme investing in the use of AR or AV. General guidance in the investigation and modelling of stimulus combination

³A projection augmented model is a type of haptic augmented reality display. It consists of a real physical model, onto which a computer image is projected to create a realistic looking object.

can be found in Ernst and Bühlhoff [35] and Jacobs [36], and with particular regard to depth in Cutting [37].

2.2.3 Augmented Virtuality

Augmented virtuality is commonly perceived to be the inverse of augmented reality, the designer intent is to dynamically integrate physical real world objects or people into the virtual world in real time [38]. This effect is usually achieved via a camera or by using digitalisation of 3D objects. The design of an AV system requires the use of approaches used to design AR and VR systems. The virtual environment should be as realistic as possible and objects in the real world need to be translated into the virtual environment using techniques borrowed from AR. Another challenging design consideration in AV is user interactivity with the system. Generally any interactivity that a designer may want to implement into an AV system is already a factor governing performance issues in AR and VR domains. For instance, there is a lack of haptic feedback in VR environments; no effective methods have yet been developed that can provide truly realistic real time haptic feedback and sensation to objects in virtual space, although progress is being made in this area. What this means is that a real object augmented into virtual space is generally lacking intractability and thus the virtual space cannot interact effectively with the real object.

Although AV as a medium for entertainment has enjoyed a reasonable degree of success in the commercial sector, namely the PS2 eyetoy [39] rather limited research applications mean that little attention has been focused towards AV development. In fact, the term AV is rarely used and is often substituted with more well known terminology such as augmented reality and mixed reality. Generally speaking AV is considered to be a subordinate class of AR and has hence, led to confusion surrounding its definition.

The scope for AV in real world applications is rather limited and relatively unexplored in comparison with AR and VR. Generally these limitations are based on the potential for products governing their application which can be directly attributed to a taxonomy of issues regarding the technological boundaries to overcome when creating AV environments. Since it is not possible to render true photo realism in VR and equally it is not possible to have absolute registration accuracy in AR, merging both these areas of technology to create compelling environments in AV is expensive both fiscally and

computationally. For now at least popular opinion has seen VR and AR outgrow AV in terms of application, development and research. It is possible that when VR and AR technology gain a stronger foothold in the mass market, technological advancements made in their field will be borrowed to AV, as was the case with the "Magic Meeting" AR system developed by Regenbrecht et al. [40] which was later lent to the development of an AV conferencing system named cAR/PE! [41]. The drawbacks currently associated with AV development make it extremely difficult to develop compelling environments. With AV development still so premature any current studies into human performance in this field could be reasonably dismissed in the future when sufficient advancements are made such to make this medium practically employable and affordable.

2.3 Display Systems

AR systems generally employ HUDs or HMDs as display mediums. HMDs are adopted throughout the research and are of particular relevance as background reading.

The foremost and most important design consideration to make when building an AR system is how to accomplish the combining of the real and virtual world. This is primarily achieved through the use of a head mounted display device, however. AR systems may utilise hand held mobile display technology as a display choice. This is usually because of the cost and usability restrictions imposed by affordable HMD technology. Strictly speaking this diminishes the immersive human element that should be inherent in any AR system and therefore can be regarded as having little or no effect on a users perception of reality. For this reason head-down approaches are reviewed in this section but will not be considered as a design choice for the work done in this thesis.

2.3.1 Head-down Displays

Head-down displays (often referred to as handheld displays) provide an attractive way to present mobile augmented reality to the user. Most commonly implementations of handheld AR utilise personal digital assistants (PDA). Handheld devices or PDAs are more compact and easier to transport than portable PCs and nicer to handle than

bulky head mounted displays (HMDs). Examples of rapid deployment of AR interfaces include Invisible Trains (see figure 2.2), a multi user game in which users control virtual trains on a real wooden miniature railroad track [42]; The Going Out System uses both camera tracking and global positioning to augment buildings with virtual information [43]. These studies suggest that for many AR applications a handheld display is more useful than a HMD since it can be viewed by multiple users. Intuitive user interfaces can be designed to help interact with the visual on-screen information and human interaction issues such as motion sickness are avoided.

There are nevertheless a number of challenges inherent in implementing this technology, namely the lack of graphics hardware support and limited processing power. One method of resolving such issues is to use the PDA primarily as a display device and run processing operations on a server that wirelessly transmits the processed information to the PDA as was demonstrated by Beier et al. [44] and improved upon by Pasma et al. [45]. Another issue with the PDA as a display device is the limited field of view a typically 5 by 3 inch screen provides. However, this could also work as an advantage in terms of the processing overhead for the system and thus improve registration accuracy of the system.

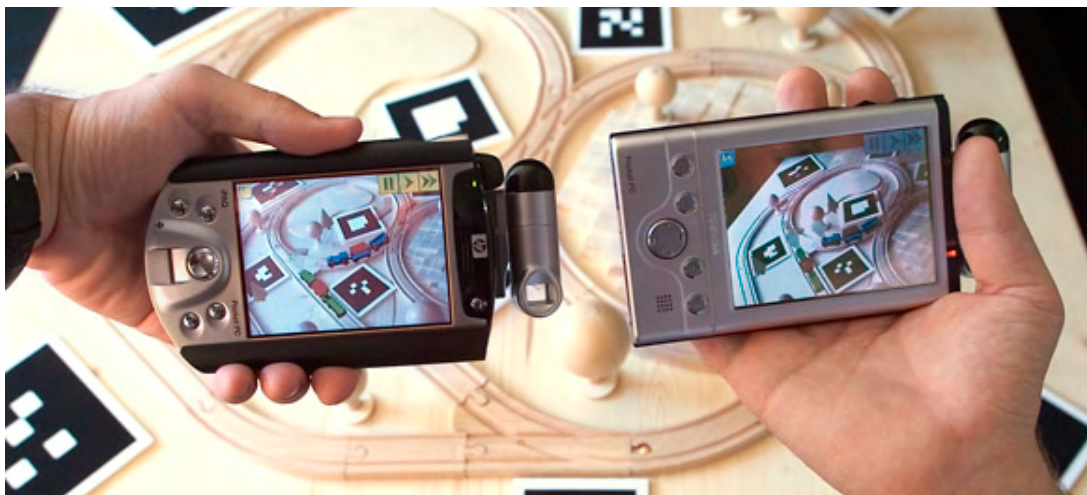


FIGURE 2.2: Invisible Trains Game Using Handheld Displays, reproduced from Wagner, 2005

The contrast between display platforms, HMD and Handheld is maintained by two critical task specific attention variables: the first being the costs to focused attention. This is related to the clutter of overlapping imagery which occurs when information is presented head-up so that it is superimposed on the outside scene creating a cluttered

view. The second variable is concerned with the cost to divided attention, or information access. When information is presented head-down the operator (pilot or driver) must now scan between the display and the outside world [46, 47].

2.3.2 Head Mounted Displays

In VR the design choice for a HMD is a simple case of sourcing the best immersive HMD available. Since the environment is wholly synthetic there is no need to use a display that shares visibility with the surrounding real environment. In AR however, there are two principal manifestations of display systems that utilise HMDs as their primary device for visual interfacing; Monitor-based or WoW (window on the world) (see figure 2.3) refer to a display system where the user views the world in real time via a capture-display device. This usually involves a webcam which feeds the images to small LCD screens placed close to the eyes; Optical see-through HMDs require that the user has the ability to see-through the display medium to the surrounding real world. This is achieved by placing partially transmissive optical combiners in front of the eyes. Each system has particular advantages and disadvantages pertained by their conceptual design and it is necessary to consider the trade-offs between the two before opting to implement one as part of an AR system. It should also be noted that non-immersive HMDs may be monocular or binocular (see figure 2.3), distinctions between the these must also be considered when choosing to design an optical based system, [48].

Optical approaches are simpler and cheaper than WoW. The world is seen directly through the optical combiners, meaning there is only one video stream to be concerned with. In see-through systems the only information available about the users head location comes from a head tracker. The WoW approach however, provides additional information via streaming digitised video image computer generated images for which virtual imagery can be analogically or digitally overlaid onto. This is achieved by using a technique called chroma-keying often observed in movies with video special effects. The background of the computer graphic image is set to a specific color, say green, which none of the virtual objects use. Then the combining step replaces all green areas with the corresponding parts from the video of the real world. This has the effect of superimposing the virtual objects over the real world. The computational expense of executing chroma-keying techniques in real time can lead to poor performance. This can

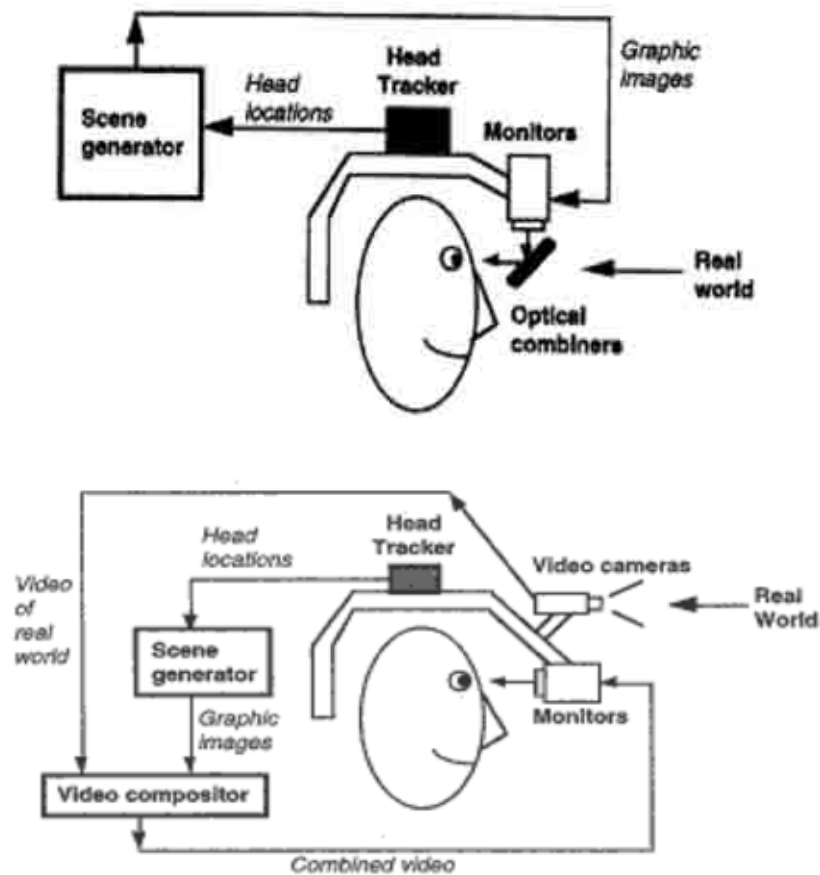


FIGURE 2.3: Conceptual Diagrams for Optical See-through and Video Based HMDs. Reproduced from Azuma [49].

present a problem of temporal distortion when real and virtual images are not properly synchronised. Thus, using post-processing computation on video streams results in time delays of a few nanoseconds with optical systems to tens of milliseconds with WoW systems.

Angular resolution describes the resolving power of any image forming device. This includes; microscopes, telescopes, optical displays, cameras or the naked eye. Unaided the human eye can resolve to the power of around 100 micro metres. Though the eye receives data from a broad range of about 200 degrees, the acuity over most of the range is poor. High resolution images must fall on the fovea and as a result acute vision is limited to around 15 degrees. The fovea is located in the centre of the retina and is 2.5 - 3 mm in diameter. Since the fovea provides the most detailed and colourful information, the eye ball is constantly moving, enabling light from the object of primary interest to fall on this region. The actual perception of a scene is constructed by the eye-brain

system in a continuous analysis of the time-varying retinal image [50]. The significance of this fact is that all facets of human perception depend heavily on movement.

TABLE 2.1: A Table to Show the performance characteristics of Low-High end HMD devices

Manufacturer	Model Name	Horizontal Res	Vertical Res	Horiz FOV	Angular Res	Cost (USD)
Cybermind	Visette 45 SXGA	1280	1024	36	1.69	\$9,649
iO Display Systems	i-Glasses 3D Pro	800	600	21	1.58	\$949
Daeyang	SXGA DH-4400VP	800	600	31.2	1.935	\$2000
Micro Optical	SV-6 PC	800	600	16	1.5	\$3,000
NVIS	nVisorST	1280	1024	50	2.2	\$34,800
SaabTech	Saab Advisor 150	1280	1024	36.8	1.73	\$95,000
Sony	Glasstron PLM-700	800	225	30	2.2	\$1,400

Currently there is no technology available that has the resolving power of the fovea in the human eye [51]. Since optical see-through technology requires that only virtual images are shown at the screens resolution the users view of the real world is not degraded. With WoW systems however, the resolution of the real world is also reduced. Table 2.1 shows examples of HMD devices and their relative properties. Consideration of the compromise these properties may introduce to human visual acuity, contrast sensitivity and depth perception (see Chapter 3) is therefore necessary for the design and utility of any AR system development.

2.4 Tracking Technology - Tracking with Hardware

Tracking hardware provide vital information about a users movement and behaviour in real world space. Both magnetic and inertial trackers are utilised such to monitor and analyse human behaviour during trial experiments.

Conventional AR system approaches offer no obvious way of getting information from the real world, thus it is up to the designer of the AR system to find a suitable method of tracking the whereabouts of the user, process this data and ultimately produce virtual images that will in real time augment correctly on the real world. Tracking the various degrees of freedom facilitated by movement on foot, head rotation and eye movement requires that various technologies be employed. Ultrasonic, magnetic and optical technologies can be employed to achieve very accurate positional data in room-sized areas [52] [53]. Head tracking is achieved using accelerometers and gyros and finally corneal

reflection can be used to track the x-y position of the eyes gaze every 1/60th second [54]. This presents a great challenge in designing the AR system since the tracking method employed directly impacts the way in which the end user will interact with the information presented to them. In light of this it is fair to conclude that the tracking method utilised may have a direct affect on human performance in virtual and mixed reality environments.

Tracking the complete pose of a user on every axis of freedom in real space can be achieved by using a combination of technologies. Often tracking is solved on a two tier approach. Firstly the orientation plane of reference pertaining to the users head in real space is defined by three elements Pitch, Roll and Yaw. Secondly the co-ordinate frame of reference pertaining to the users position in real space is defined by three elements X, Y, Z (see figure 2.4).

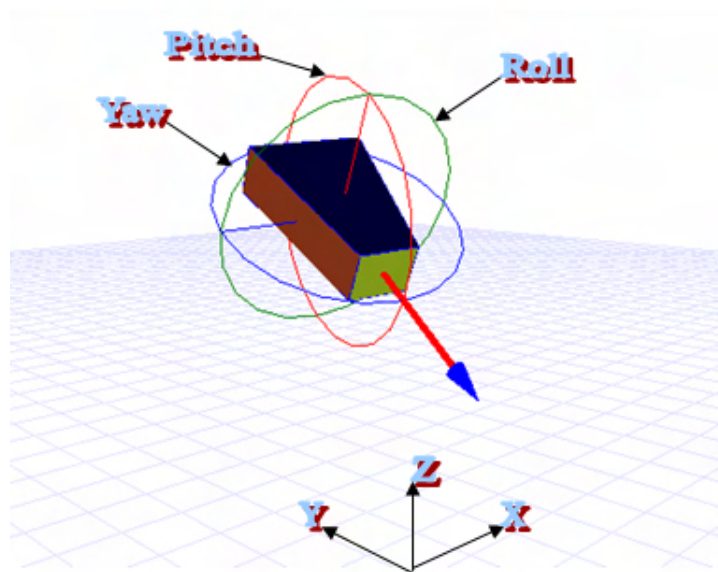


FIGURE 2.4: Visualisation Plane showing All 6 Degrees of Freedom in Movement

2.4.1 Lidar

A LIDAR (Light Detection and Ranging) also known as Laser Range Finder is a device which uses a laser beam in order to determine the distance to an opaque object. The technology works by sending a narrow beam of light at the target and measures the time it takes to return to the sender, described simply by the equation $\text{Distance} = \text{Speed} \times \text{Time}$.

The accuracy of a LIDAR instrument is determined by the brevity of the laser pulses it emits and also the speed of the receiver. Thus, it follows that a LIDAR which uses very short sharp laser pulses and has a fast detector should be capable of estimating an objects distance to within a few centimetres. Where laser range suffers is on surfaces that will not reflect the laser beam [55]. The ability to measure a users distance from real world objects can be used to alter scaling of virtual information and simulate depth in the MR environment. In two systems created by Grimson et al. [56] the problem of registering virtual objects over live video is solved as a pose estimation problem. By tracking feature points in the video image these systems invert the projection operation performed by the camera and estimate the camera's parameters. Using a LIDAR the Euclidean 3D location of feature points can be found and therefore the camera parameters can be estimated in a Euclidean frame.

2.4.2 Stereo Camera

Using two cameras set at a short fixed distance from each other it is possible to accurately estimate the distance to an object. To improve accuracy and usability of the technology a powerful laser is positioned precisely between the two camera lenses. Upon taking a photo the laser beam is focused on the object of interest then when the photos from both cameras are analysed, distance to that object can be estimated by finding the pixel difference of the laser reflection on the object surface on each picture [57]. Another effective way of implementing this would be to calculate the distance between spatially identical edge pixels on each image. Rigging up such a system manually is very error prone as any compromise to the camera lens distance influences the range finding results of the system considerably. This approach is often employed for use in feature based stereo AR systems which compute sparse depth measures based on correspondence of high level features detected independently in each image.

2.4.3 Ultrasonic

Measuring the distance to an object using sound waves is a popular, simple and cost effective method. Distance is found by measuring the time it takes for an ultrasonic sound signal to be sent and received. This technology is generally employed for estimating distance to large flat surfaces such as a wall or window. However, accuracy suffers

when trying to negotiate the distance to smaller objects because the sound wave may only be partially reflected [12]. Co-ordinate frames of reference can be achieved using an array of ultrasonic receivers placed around a closed space. For instance, if an ultrasonic emitter is placed on the user. A users position can then be triangulated in real time based on their distance from three or more sensors [58].

2.4.4 Magnetic Trackers

Magnetic trackers are a popular and cost effective way to record the pose of an object in all six degrees of freedom. A magnetic tracking system consists of a transmitter that has three coils on orthogonal axes. A current is passed through each coil. The sensor consists of a similar set of three coils. Depending on the system, varying signal strengths or time multiplexing is used so that each of the three magnetic fields can be isolated. This gives enough information to determine the difference between current loss due to rotation and movement. In the past magnetic tracking systems suffered greatly from disturbance particularly in the presence of metal and electric equipment but have remained popular because they are robust and place minimal constraints on user motion [59]. AC magnetic tracking technology offers many design variables to reduce interference, high conductivity materials like copper, aluminium, brass and some steels definitely can sustain eddy currents for distortion. However, this is insignificant unless the material is very close to either the source or the sensor or it is large enough to sustain appreciable eddy currents. The polhemus patriot⁴ documentation for example provides a list of metals that have little or no distortion effect. The presence of magnets also is of no consequence unless they also happen to be highly conductive (small, strong ferrite magnets have been tested with no effect) even when brought very near to the sensor.

Compared to inertial trackers the source of a magnetic tracker is the co-ordinate reference so there is no initialisation, drift correction or subtracting off movements with a second system. Magnetics have no line-of-sight problems like optical and acoustic approaches, allowing embedding of a sensor anywhere in the HMD or visual display panel. Another point is that AC magnetic trackers automatically yield from a single sensor both position and orientation on every sample, unlike other techniques that yield only position or

⁴http://www.polhemus.com/?page=Motion_Patriot

velocity and must derive position or orientation from multiple readings. Despite efforts however, magnetic trackers are subject to inaccuracy in their reports, which in the case of magnetic based tracking systems, is often caused by local magnetic interference. In AR environments, inaccurate reports lead to mis-registered synthetic elements. To reduce this error, data from the magnetic tracker should be corrected before being passed to the image generator [60]. Despite error correction efforts inaccuracy continues to be a problem. This coupled with high hardware and environment implementation costs, ultimately leads to vision based tracking methods generally being favoured over magnetic implementations.

2.4.5 Inertial Trackers

Inertial trackers use gyroscopes and accelerometers to measure changes in angular velocity and linear acceleration. The pose of an object can be accurately measured using an accelerometer to measure movement in the pitch and roll domains of orientation. However, accelerometers cannot be used to measure yaw, since the axis upon which it is measured is perpendicular to the force of gravity and is therefore unaffected by it. Accelerometers can therefore only be used to calculate pose in two of the three orientation domains yet it is possible to stabilise accelerometer measurement in the yaw domain using gyroscopes. This is achieved using accelerometers on a gimbaled gyro-stabilised platform. The gimbals are a set of three rings, each with a pair of bearings initially at right angles. They let the platform twist in any rotational axis. There are two gyroscopes (usually) on the platform. Two gyroscopes are used to cancel gyroscopic precession⁵ by mounting a pair of gyroscopes (of the same rotational inertia and spinning at the same speed) at right angles the precessions are cancelled, and the platform will resist twisting. Using this system pose in the roll, pitch and yaw angles can be measured directly at the bearings of the gimbals. Relatively simple electronic circuits can be used to add up the linear accelerations, because the directions of the linear accelerometers do not change. The big disadvantage of this scheme is that it uses many expensive precision mechanical

⁵Precession describes a phenomenon in which a small rotation in a plane perpendicular to the plane in which the gyroscopes symmetrical mass is spinning resulting in a tendency for the gyroscope to twist at right angles to the input force.

parts. It also has moving parts that can wear out or jam, and is vulnerable to gimbal lock⁶

2.4.5.1 Accelerometer

An accelerometer is a sensor that converts mechanical vibrations into electronic signals that are proportional to the vibratory acceleration 2.5. Once acceleration is obtained tilt can be determined. This occurs because changing tilt along the sensitive axis changes the acceleration vector. When a tilt force is exerted the accelerometer outputs a voltage (V_{out}) which allows us to determine the angle of θ , movement:

$$\theta = \arcsin \frac{V_{out} - V_{Og}}{1_g x Sensitivity} 1g = 9.8m/s^2$$

where V_{Og} is the zero g offset (the voltage output at zero g), *Sensitivity* is a constant measured in Volts/ g .

Conventional accelerometers based on MEMS (Micro Electro Mechanical Systems) technology work using polysilicon springs which support and control the movement of a beam. This moveable mass responds to vibrations which are measured by a differential capacitor whose output is proportional to acceleration. Acceleration is determined using the following two laws of physics:

Newtons 2nd Law: The force (F) on a mass (m) subject to acceleration is: $F = ma$

Hooke's Law: The deflection (x) of a restraining spring is proportional to the applied force (F). $F = kx$

Hence $a \propto x$

⁶Gimbal lock is the loss of one degree of freedom that occurs when the axes of two of the three gimbals needed to apply or compensate for rotations in three dimensional space are driven to the same direction.

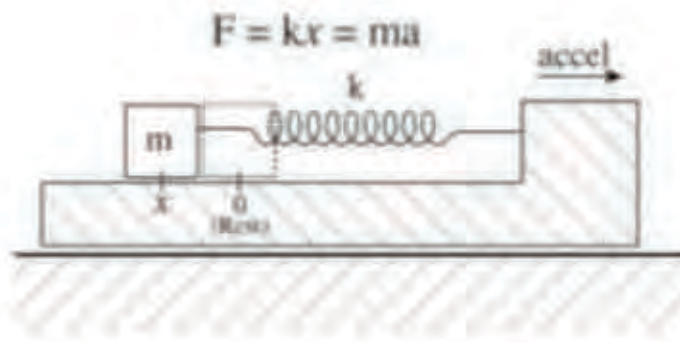


 FIGURE 2.5: Accelerometer Based on MEMS technology

2.4.5.2 Gyroscope

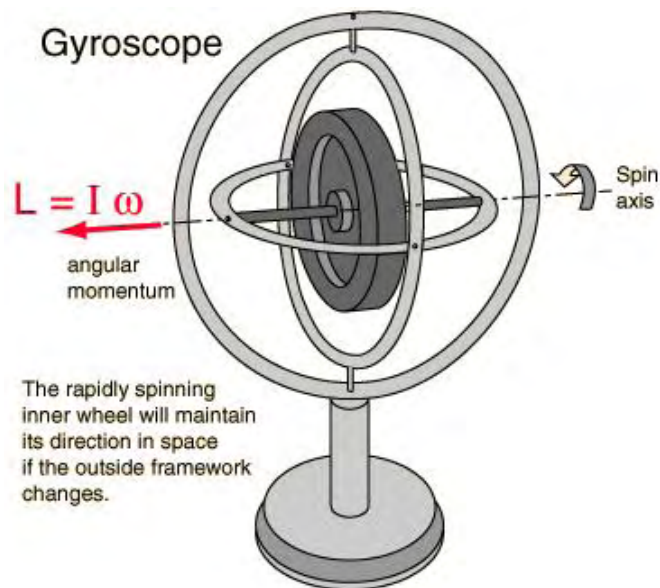


 FIGURE 2.6: Gyroscope - Symmetrical mass is mounted within a three tiered frame called a gimbal.

A gyroscope is a device consisting of a spinning symmetrical mass that is mounted so that it can spin about any direction (see figure 2.6) When this perpendicular axis is coned by a gimbal the gyroscopes behaves such that it only resists a tilting change about its own axis giving it a property that can be used to measure tilt about that axis very accurately [61]. MEMs technology has made it possible to create silicon chip based gyroscopes for angular-rate sensing at very affordable prices.

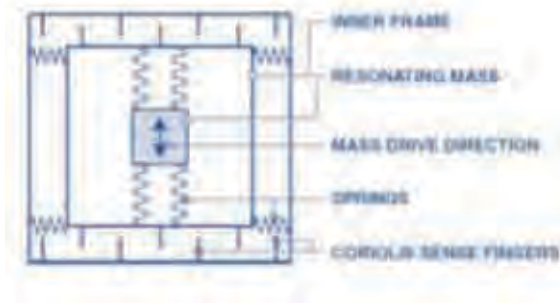


FIGURE 2.7: Gyroscope. Movement of a resonating mass is measured by coriolis sense fingers.

These gyros maybe placed anywhere on a rotating object and at any angle, so long as its sensing axis is parallel to the axis of rotation. The angular rate of change is found based on the principal of french mathematician Gaspard G. de Coriolis law of acceleration that the rate of increase in your tangential speed is caused by your radial velocity. These gyros take advantage of this effect by using a resonating mass that is micro-machined from polysilicon, tethered to a polysilicon frame so that it can only resonate in one direction [62]. Referring to figures 2.7 and 2.8 shows that when the resonating mass moves toward the outer edge of rotation it is accelerated to the right and exerts a force to the left, this force is then measured by the coriolis sense fingers which capacitively sense displacement of the frame in response to force exerted by the mass. This displacement (D) resulting from the force exerted by the mass (M) is calculated as follows:

$$D = \frac{2\Omega v M}{K}$$

where Ω is the angular rate of acceleration, K is the stiffness of the springs.



FIGURE 2.8: Gyroscope. Rotation forces the mass to move toward the outer edge when moving in the clockwise direction and moves towards the inner edge when moving in the anti-clockwise direction.

2.5 Pattern Recognition - Tracking with Software

Computer Vision techniques provide an effective, easily employable method for achieving registration and augmentation in the real world. The most commonly adopted method utilises fiducial markers and is the chosen tracking method for the AR system developed for this research

There have been many tracking systems developed that utilise vision based techniques to achieve accurate registration. In prepared environments where the system designer has sufficient control over the environment most have adopted a technique using visual fiducials (pre-designed artificial landmarks) to calibrate the environment for augmentation of virtual data (see figure 2.9). In these systems an estimate for the camera pose is identified by tracking a fiducial whose 3D co-ordinates are calculated prior to system operation [63] [64]. In order for AR systems to maintain the perception that the virtual data is part of the real world, accurate real-time computation of the camera pose relative to the coordinate system of the virtual data is crucial. The ARToolkit developed by the HIT lab at the University of Washington has designed such a system that uses markers for registration. These markers are comprised of a thick black border around a symbol. When the marker border is identified the symbol is converted into a binary image and template matched. Other techniques include the use of colours as marker points. Such a system was employed by the developers of the ARHockey System [65]. Although registration using this technology was fairly accurate ambient light in the vicinity of the system had to be very carefully controlled. It is also possible to achieve markerless pose estimation using affine region tracking techniques to track patches or a feature set in the real world to augment virtual imagery to [66], although compelling in concept this approach relies on building correspondences between 2D feature points that are trained into the system prior to use. This requires that the scene for augmentation is prepared such that the trained real world feature set is always available.

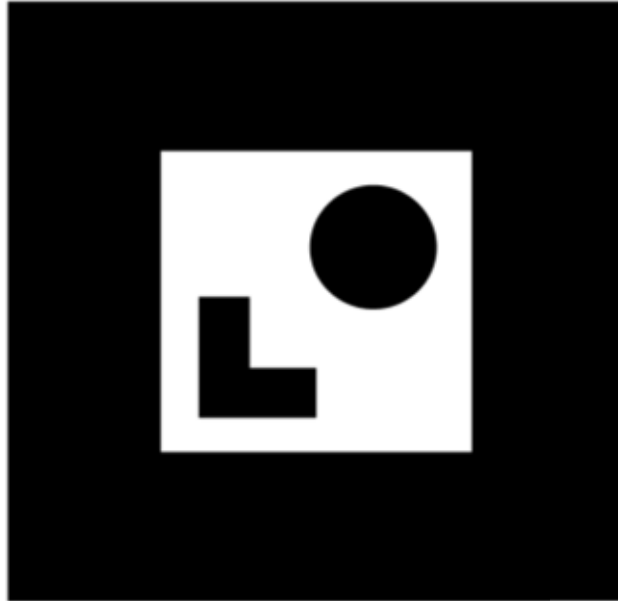


FIGURE 2.9: Sample of a typical fiducial used in marker based augmented reality applications.

2.6 Discussion

The literature in this chapter has shown that on the whole the challenges and goals facing the development of any VR or AR system share many distinct similarities

- Both focus on the disappearance of conscious and intentional interaction with an informational system and try to embellish the user with a holistically natural method of interfacing.
- Both provide the user with virtual representations of real world objects that are indecipherable from real world counterparts.
- Both track and register the natural movement of the user in the real world and replicate it (exactly) in the proposed mixed reality or virtual space.

The most salient distinction to be made between AR and VR technology is that AR is focused on enhancing human experience in the real world as opposed to immersing them in a synthetic one. Also the definitions pertaining to VR systems is clear and precise whereas problems may arise in MR situations. The work of Milgram et al. [21] (refer to 2.1), proposed the reality-virtuality (RV) continuum as a way of clarifying various

states of MR and proposed an objective distinction between reality and virtuality based on three distinct criteria:

- A first distinction between real and virtual objects by means of the following operational definitions: a. Real objects are any objects that have an actual objective existence like, for example, the computer I am using to write this document. b. Virtual objects are objects that exist in essence or effect, but not formally or actually, that is, they can also be existing objects, but they do not exist here and now. Therefore, a real object can either be observed directly or it can be sampled and resynthesized through some display device. A virtual object, instead, cannot be directly observed since it does not exist and therefore must be simulated. To this aim, a description or a model of the object is usually needed.
- A second distinction concerns the issue of image quality as an aspect of reflecting reality. On the one hand, as stated above, virtual objects cannot be directly observed nor sampled: they can only be synthesised. On the other hand, modern technology makes synthesis of extremely realistic images possible but nevertheless this does not make them any more real.
- A third distinction is made between real and virtual images. A real image is defined as any image which has some luminosity at the location at which it appears to be located. Virtual images are conversely defined as images not having luminosity where they appear. Virtual images include holograms, mirror images, and stereoscopic displays (for which both the left and right images are real images, but not the fused image). Virtual images in MR environments are transparent and as such they cannot occlude the objects located behind them.

The disparity between AR and VR and their roles will become more apparent, the causality being the third paradigm of computing that will come to dominate human computer interaction over the next few decades [11]. As pervasive and ubiquitous interfaces move to the forefront of computing technology, AR system design will be able to draw from these advances and become integrated within these context aware real world workspaces. VR however, may well have to draw on other developments in the scientific community to find new possibly more invasive methods to make translational movement in virtual worlds more realistic. The fact that VR technology is currently still

fairly immature in terms of natural human interfacing indicates that AR is probably a more suitable medium for investigating the affect on human performance when a user's view of the world is changed. Using AR to study affects on human performance will centre research wholly on alterations to human perception without introducing possible bias from non-natural interfacing methods that a VR system would impose. Despite the advantages of AR over VR with regards to natural human interaction and interfacing, there will always be a conflict between how virtual information behaves and how a user perceives and reacts to information they perceive to be unreal. Suffice to say, if the user is conscious that a scenario is not real the way they interact with it may not mirror real life effectively regardless of how realistic that experience may appear to be. Current research focus has largely been upon evaluating interaction with synthetic imagery and ignores the affects this may have on the perception of the real world itself. This leads to a possible confound in studies that have compared perception of real and virtual objects, this being that it is commonly assumed the perception of real objects is veridical, and the difference between estimations of the positions of real and synthetic imagery is used as an error measure. It may however, be the case that perception of neither stimulus is accurate owing to perceptual cross-talk between the two.

Since the advent of MR technology may inadvertently change the way in which a person deals with what they perceive it is pertinent that using AR or VR as a means to measure human performance or human ability could be disingenuous and unsubstantiated. As yet, whilst different variables have been examined in gross form by AR researchers pilot testing different prototype systems, there is a paucity of genuinely parametric research; that is, where various stimulus parameters are systematically varied at different levels of intensity/accuracy/size/contrast and so on, and their interactions measured (see Swan et al. [67] for an example paradigm). Without parametric data it is difficult to suggest how a person's perception of virtual information may differ to real information. If it does differ this could determine the utility mixed reality manifestations take on and govern how effective they are as alternatives to real world interaction experiences. This being the case the utilisation of MR environments as a medium to aid or improve task performance requires further research and evaluation. For the purposes of this research, AR could provide a useful medium of complementing conventional environments used for training crime scene investigators. These conventional or 'mock' crime scenes could benefit from AR by for example utilising a fast scene re-configuration for implementation

of differing modals of crime, an accurate and instant measurement of user performance and an on-demand position based guidance narrative.

Chapter 3

Synthetically Enhanced Reality - Human Visual Perception

3.1 Introduction

The technology used to create AR and VR systems varies dramatically in both design, software development techniques and programming approaches. Technology and system design choices will ultimately affect a viewers interaction and cause them to perceive the world differently. The research in this chapter hopes to ascertain what considerations and choices should be made when developing an AR system. Each facet of the system design will affect the final system composition and have ramifications in regards to the effect on human performance.

3.2 Human Perception

The underlying cause affecting human performance relates to a number of factors pertaining to perception of the environment. However since movement is the most salient measurable factor affecting performance, particular attention should be paid to related literature.

It has been shown that difficulty lies in the algorithmic nature of perception itself [68]. The real world exists in three dimensions but the sensor that encodes light information,

the eye, is two-dimensional, meaning the brain is faced with the difficult computational problem of inverse optics (recovering the structure of the 3D visual scene from the information available in the 2D optical array). As Robert A. Jacobs notes in a statistical analysis of cue reliability:

Every visual cue is ambiguous. There are many reasons underlying this ambiguity, including physical factors, such as atmospheric or optical blurring, and biological factors, such as noise inherent to human nervous systems. Therefore, there is no correct interpretation of a cue.... [36]

As a result of these inherent difficulties the visual system tends toward being probabilistic in its interpretation of cues. Richard Gregory has referred to perceptions as hypotheses for this reason:

Perception is not determined simply by the available stimulus patterns; rather it is a dynamic searching for the best interpretation of the available data...The senses do not give us a picture of the world directly; rather they provide evidence for checking hypotheses about what lies before us. Indeed, we may say that a perceived object is a hypothesis... [23].

Plausible outcomes for the combination of synthetic depth cues with the real world are as follows:

- Synthetic imagery matches real world, meaning the perception of both are accurate.
- Synthetic imagery is discordant with real world and synthetic imagery is rejected as false.
- Cues are in conflict.
- Cues are combined/averaged leading to a distortion in perception of real and/or synthetic imagery.
- Cue conflict cannot be resolved leading to perceptual rivalry, typically resulting in a bistable percept that changes over time (presumably due to the action of noise within the perceptual system tipping the balance) or changes due to other subject variables (such as cognitive biases and expectations).

One of way of thinking about this is by considering perception of depth as analogous to solving a jigsaw puzzle in which information from different cues are the jigsaw pieces. The placement of a piece within the jigsaw is multiply constrained, which is to say the placement of a piece will be limited by how other pieces are already put together, and under normal circumstances this is helpful in determining where an otherwise ambiguous piece should go. As we are able to put together more and more pieces our confidence that we have done things correctly grows. The challenge of creating AR stimuli then is to create jigsaw pieces that are complimentary to those already present in the natural world. It may be the case that these new pieces cannot be fitted in anywhere (rejection), that these new pieces cause us to construct the jigsaw wrongly (distortion) or in rare cases the additional pieces will happen to allow the jigsaw puzzle to be solved in numerous ways without any clear indication of which solution is correct (bistability).

This theory of perception could be used to speculate how the technology employed to design an AR system may affect human performance. The most salient deduction to be made here is that choices in technology, the implementation of system design and the encumbrance this places on human interaction with the real world will have a dramatic affect on human perception and thus performance. Form factor variations in AR displays for instance have a direct impact on the behaviour of natural human interaction in the real world such as; divided focus of attention (head down displays), sensitivity to contrast (see through HMDs), visual fatigue and latency (monitor based HMDs).

3.2.1 Body-Based Movement

The way in which an environment is perceived is intrinsically linked to how humans are able to move around it. People navigate their environments using two frames of reference: a frame which is defined with reference to the environment itself, e.g., Ego-centric which is defined with reference to the person, in terms of orientation of the eyes and body, and which they would assume to alter under their own volition and which allows them to scale the environment, and Exo-centric which is defined in terms of Cartesian co-ordinates, so that objects typically remain in their fixed positions (unless moved), and so that rotation of the person would be relative to a fixed world. A particular difficulty, when moving in virtual environments, lies in the lack of concordance between motion cues and the sensations of movement that the person normally feels in the real

world which makes it difficult for users to compute their heading, this results in users getting lost easily in VR systems and experiencing feelings of discomfort [69]. Riecke [70] showed that navigation problems persist even when high-quality video projection systems are used and when advance information about turning angles was provided. This suggests that the problems are not solely related to the sophistication of the technology but relate to the manner in which the users attempt to combine sensory cues to interpret movement. Earlier studies by Ruddle et al. [71] and Bowman and Hodges [72] stress the importance of allowing the user to look around whilst moving. The following notation (reproduced from Ruddle and Jones [73]) describes how movement can be facilitated in virtual environments:

The challenge in designing a VR system is how to facilitate human movement in a synthetic environment. VR systems facilitate movement through the coupling of three types of dependent/independent directional movement in the horizontal plane:

- H_b - orientation of the body (direction in which body is facing).
- H_v - direction of view (which way person is looking).
- H_t - direction of travel (such as forwards, backwards, sideways).

Thus real world movement is expressed as:

$H_b = H_v = H_t$ where $=$ is used to indicate that directions can be varied independently.

Such has been but is rarely facilitated by VR systems [74]. Traditionally most virtual environment (VE) interfaces used view-direction travel (travel where you look) $H_b = H_v = H_t$. However it has become increasingly popular to decouple H_b and H_v , hence $H_t = H_b = H_v$ thus allowing for independent looking around whilst moving. Neither of these methods support independent variance of H_t

In virtual environments, the decoupling of different frames of reference can be technically challenging and (more importantly) potentially disconcerting for the person. Some older VR systems (and, by extension, video games) assume that all aspects of movement are tightly coupled, so one moves in the direction that one is facing, although contemporary systems tend to support a combination of moving in the direction of the body's orientation but allow movement of the direction of view. These types of movement can

be implemented in both immersive and desktop VR using a variety of different devices, and, clearly, the device that is chosen can have a substantial effect on user performance. Interface devices can be characterised by factors such as the number of degrees of freedom (DOFs) that can be simultaneously varied, the order of control that is used (zero order (position), first-order (velocity) or second-order (acceleration)), and the range of values that are measured (Baber, 1997). In desktop virtual environments, a pointing device, such as mouse, joystick or cursor keys, remain the most common interface devices, and these can be used to implement either view- or body-direction movement [73]. In immersive virtual environments, sensors can be used to track the movement of parts of a user's body and the user typically controls movement speed by pressing, or holding down, buttons. The use of body-mounted sensors, e.g., on head or waist, allows gross movement to be tracked. Studies by Ruddle and Lessels [75] have demonstrated that interaction in virtual environments, on tasks that require movement in space, tend to be superior when body movement is supported rather than the use of a pointing device (when the person is usually static and controlling the device while seated). The wealth of research relating to body based movement in wholly synthetic environments is helpful in predicting how effective implementations of AR can be expected to be. On the whole very little research has been done that directly correlates the effects that AR has on human performance as it pertains to body based movement. A study by Cao and Milgram [76] shows how the addition of AR to enhance a scene with additional information about location and direction improves user accuracy in a blind, non rigid environment. However, this is principally a study in navigation. There is most certainly a paucity of research that links body based freedom of movement to performance in AR.

AR inherently has the ability to support a complete range of motion. Both VR and AR present a similar challenge for the designer in that tracking is key to instilling the user with a sense of immersive disbelief. In VR accurate tracking of human movement will create an environment that feels natural and non-invasive. In AR such will allow virtual objects to appear naturally in the real world. Unfortunately, using current adaptations and implementations of technology it is not possible to register virtual information into the real world infallibly and it is these subtle shortcomings in registration that affect the way in which users interact and behave in an environment.

A frequently researched phenomenon in VR is underestimation of distance. An attempt was made in a study by Willemsen et al. [77] to compensate for this perversion of depth

perception. which causes even photorealistic virtual environments to appear compressed when compared to visually identical real environments (viewed through identical optics). In the study they manipulate stereo viewing conditions in a head mounted display and show the effects of using both measured and fixed inter-pupillary distances, as well as binocular and monocular viewing of graphics, on absolute distance judgments. The results indicate that the amount of compression of distance judgments is unaffected by these manipulations. The equivalent performance with stereo, binocular, and monocular viewing suggests that the limitations on the presentation of stereo imagery that are inherent in head mounted displays are likely not the source of distance compression reported in previous virtual environment studies. This suggestion that human perception of distance is affected by something so subtle that current research has been unable to determine its cause begs the question in relation to AR and its affect on human performance. AR can only be helpful as a training tool to aid human performance in real world tasks if human behavior is clearly correlated in both environments. Otherwise, human performance in an AR environment will not be a clear indication of performance in an equivalent real life scene.

3.2.2 Depth Perception

The relative importance and contribution made by different depth perception cues varies with distances involved. As distance increases pictorial rather than physiological cues become dominant [37]. For most AR observations, users operate at short ranges, up to 50 metres. Here dominant cues include motion parallax, ocular accommodation, convergence and binocular disparity. There are some difficulties in predicting how various cues may interact at the perceptual level (although the chart given by Cutting [37], suggests which cues will be available and prominent at different contrast levels). This is particularly problematic for AR/AV applications where we might wish to introduce our own depth cues that may be at variance with those in the real world (either deliberately or because of problems in implementation).

3.2.2.1 Depth Perception in AR

Studies that have looked specifically at depth perception suggest that there are difficulties in producing virtual imagery that contain accurate depth cues. Rolland et al. [78]

found that mixed experimental design where virtual objects were presented alongside real objects the virtual objects tended to be perceived as located further away than real objects. Rolland et al. later examined in a 2002 study [52], occlusion of real objects by virtual objects and found that even when the virtual object was presented in front of the real object, it was perceived as standing behind. However, it was also noted that observers were better able to determine the correct location of the virtual object when allowed to adjust the depth of the virtual object. Contrastingly, Ellis and Menges [79] using different technology and software found that virtual objects tended to be perceived as nearer than they really were, although this effect was highly variable with regard to participant factors (age and ease of varying ocular accommodation) and error could be reduced by placing virtual backgrounds or textures. This suggests that incongruence between the virtual and real can lead to misperception of depth which impacts upon depth judgements of virtual objects rather than real world objects.

A relative paucity in research related to depth perception with regards to AR systems means design choice for AR systems depends mostly on task requirement. Monocular see through displays may cause rivalry between perceived depth of virtual versus real information, where binocular see through displays may eradicate perceived issues with real objects, disparity in the virtual domain may introduced by way of binocular rivalry.

3.3 Vision

The world exists in three dimensions but can only be visualised in two. It is therefore important to understand how these underlying factors such as depth and visual acuity affect vision and consequently perception of the environment.

3.3.1 Field of View

The field of view is the angular extent of the observable world that is seen at any given moment in time. Vision through an optical system may be obstructed and limited by the veiling luminance of the display and the supporting structure (body) of the HMD. Figures 3.1 and 3.2 depict the area between the display and the body. This section

(called clearance) can obstruct the view of an object in a particular direction, this obstruction called a scotoma is more prominent in binocular HMDs. Monocular HMDs are an essentially open environment with little or no constraint on the users field of view [80]. There are however, safety issues with monocular HMDs concerning the ability to divide attention between the display and the real environment.

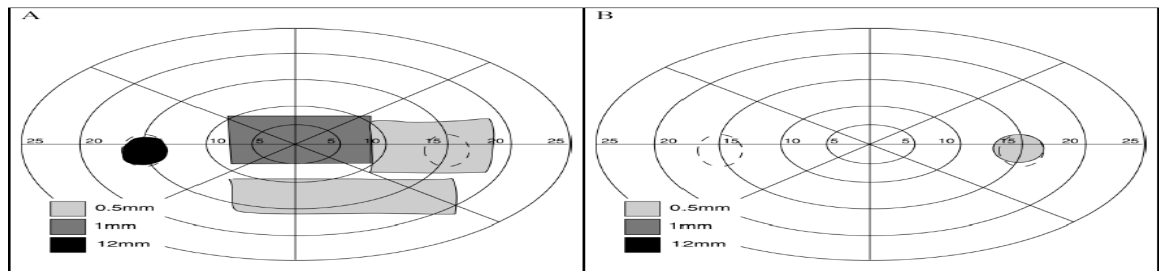


FIGURE 3.1: Central Visual Fields with MicroOptical Eyeglasses. - (A) The monocular visual field (left eye) showed the physiological blind spot and relative scotomata from the display and from the optics of the device. (B) There was a relative scotoma in the binocular visual field due to the overlap of an optics scotoma and the physiological blind spot of the other (right) eye. Reproduced from Woods et al. [80].

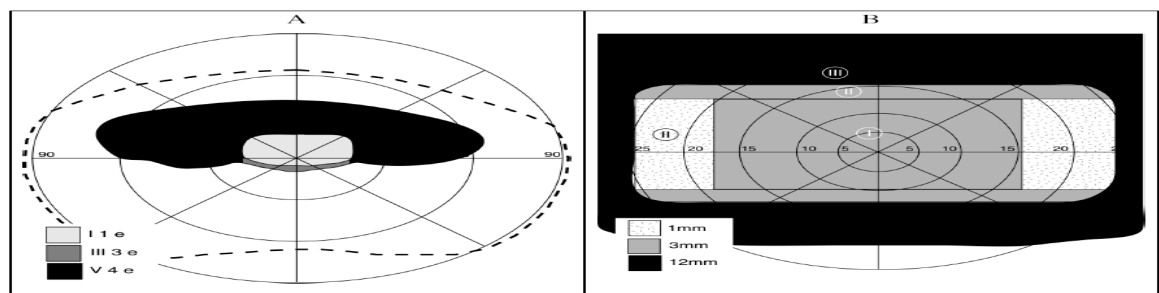


FIGURE 3.2: Binocular Visual Field with Sony Glasstron PLM-50 - (A) to $\pm 90^\circ$ with a Goldmann perimeter; and (B) with blue screen on to, $\pm 90^\circ$ with the B&L Auto-Plot perimeter. The areas of interest are (I) the relative scotoma caused by the display, 3 mm targets were not seen in this region; (IIa) a relative scotoma in the clearance on each side of the display, 1 mm targets were not visible; (IIb) a relative scotoma in the clearance above and below the display, 3 mm targets were not visible; and (III) an absolute scotoma caused by the HMD body that extended more than 60° to either side. The extent of the normal binocular visual field shown as the dashed line in panel A (V4e target). Reproduced from Woods et al. [80].

Inherent in any video-based MR system is the constraint on visual awareness. Forcing the user to see the world through cameras disassociates the users autonomy with the real world. The primary cause of disassociation is the users view of the world is constrained to a limited field of view and prohibited by a fixed degree of freedom that only allows for movement of the head and body whilst ignoring eye movement. The field of view values in most HMD devices is given to represent the FOV of the device with a 100%

stereoscopic overlap configuration. At least 20% overlap is required to satisfy the human visual system and create a satisfactory sense of depth. The human eye has a total FOV of 160° to 208° , about 140° or so for each eye, yet 120° - 180° is attributed to binocular vision. Ignoring monocular see through displays, many HMD manufacturers use stereoscopic overlap techniques to achieve a wider effective FOV (see figure 3.3). The disadvantage of this technique is that stereoscopy is lost in the monoscopic region. Ultimately the effect this has on the user and their performance depends heavily on how they interact with the device. Since the stereoscopic central region does not account for human eye movement, users may experience disturbing effects referred to as binocular rivalry.



FIGURE 3.3: Monoscopic and Stereoscopic Overlap Region for Human Eyes.

3.3.2 Binocular Rivalry

Binocular Rivalry occurs when each eye is presented with a different image (see 3.4). With the transparent monocular configuration of an optical HMD system one eye views the real world whilst the other eye views virtual images superimposed onto the real world [81]. This can cause periods of monocular dominance which in turn is not predictable because it is not consciously controllable. Users may react to this phenomenon by closing their eyes and may experience dizziness from long term exposure to the system. Optical HMDs are constructed such that the virtual image appears at a fixed focal distance from the eye(s). Typically this is one to two metres. However real world imagery may be at

any focal distance. A real world objects focal distance may differ to that of the virtual object or text we wish to augment to it, focus therefore can only be given to each object in turn which may be discomforting and disorientating to the user. Methods have been developed to combat the effects of binocular rivalry to perception. A study by Yamazoe et al. [82] showed that by inducing optokinetic nystagmus (OKN)¹ it is possible to reduce binocular rivalry, however to design a system that can track and then synchronise the movement of the eyes with a camera lens is highly complex and costly.

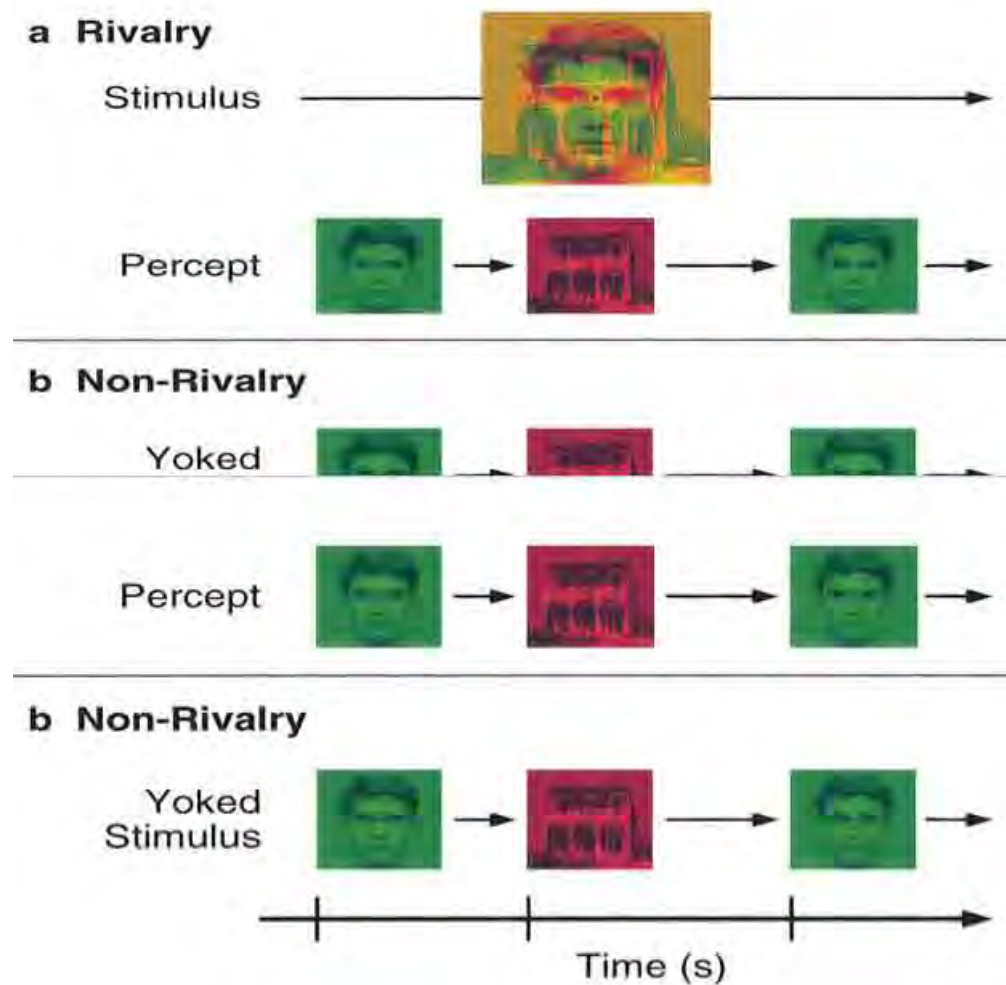


FIGURE 3.4: Experimental Design and Stimuli (a) Ambiguous face/house stimulus used in rivalry scans. When viewed through red and green filter glasses, only the face could be seen through one eye and only the house through the other eye. This led to vigorous binocular rivalry as indicated by reported alternations between a face percept and house percept (typically every few seconds). (b) A timeline illustrating how non-rivalrous monocular images of either face or house alone using the same temporal sequence derived from the perceptual report of a previous rivalry scan. Reproduced from Laramee and Ware [83].

¹The optokinetic reflex allows the eye to follow objects in motion when the head remains stationary (e.g. observing individual telephone poles on the side of the road as one travels by them in a car)

3.3.3 Eye Offset

In most WoW systems the cameras are not located exactly where the user's eyes are, thus creating an offset between the camera and the users eyes. In addition the distance between the cameras may differ to the users Inter-Pupillary Distance (IPD). This can cause difficulties with both registration and user orientation because of the inherited displacements between what the user sees and what they expect to see. This problem can be avoided using mirrors to create optical paths that mimic the optical paths of the eyes however. This adds complexity to the design of the system. For optical see-through systems offset is generally not a difficult design problem. While the users eyes can rotate with respect to the fixed position of the HMD, the resulting errors are very small.

3.3.4 Visual Acuity

It is important when designing AR systems to know that users will be able to resolve presented imagery, there is clearly no value to a system where stimuli cannot be properly seen or interpreted. In terms of absolute detection of stimuli or minimum perceptibility, estimates derived from asking participants to detect the presence of dark discs or squares against a bright background suggest that participants can detect stimuli of around 0.25 to 0.5 minutes of an arc [84]. One immediate difficulty this proposes is how clear the resolve of the displays in use will be to the user. The resolve of a display compared with the real perceived visual acuity can be described by a displays angular resolution. The angular resolution of a display represents a normalised measure taken after the image has travelled through all the optics of the device. This is a critical factor because no matter how the optics rearrange the image this measurement gives a clear indication to what the user actually sees. Referring to chapter 2 table 2.1 shows that high cost HMDs achieve a maximum total pixel resolution of $1280 \times 1024 = 1,310,720$ pixels. This falls far short of the human eyes resolving power of about 0.3 minutes of an arc. Human performance in this respect may well be affected by the fact that information viewed through currently available optics cannot match the perceived visual acuity of objects viewed in the real world. This is also true of the camera capturing a real scene. To exemplify just how

acute the human visual system is consider this calculated approximation of a view that is 90° by 90° , such as looking through an open window at a scene².

The number of pixels will be:

$$90^\circ \times 60' \times \frac{1}{0.3} \times 90 \times 60 \times \frac{1}{0.3} = 324,000,000 \text{ pixels} (324 \text{ megapixels}).$$

Note 60 arc-minutes per degree is the definition of degrees and arc-minutes. $\frac{1}{0.3}$ comes from the definition of visual acuity. Visual acuity is defined as the resolution of the eye in units of $\frac{1}{\text{arc-minutes}}$ thus a higher value results in better acuity. Taking acuity = 1.7, the eye resolves $\frac{1}{1.7}$ arc-minutes = 0.59 arc minutes. Since two line pairs are required to resolve anything, such as two stars close together. $\frac{0.59}{2} = 0.3$ (rounded value).

However, as discussed earlier at any one moment humans do not perceive this many pixels. The eye moves around the scene to see and the brain composes all the detail in the scene. The human eye actually sees a field of view, larger than 90° , it is closer to 180° . Using a conservative 120° for the field of view. An approximation for human visual acuity is:

$$\frac{120 \times 120 \times 60 \times 60}{0.3 \times 0.3} = 576 \text{ megapixels}.$$

There are issues with this rather crude approximation of the capability of the human eye. The brain composes the image on to the fovea which has 30000 cones. The actual perception of a scene is constructed by the eye-brain system in a continuous analysis of the time-varying retinal image. Therefore, it is wholly likely that even a 576 mega pixel image would not be perceived as real based on this principle because of so many other factors such as the linear light reflectivity of the scene and the static non-variance in the light from the 2D source. Despite this, the approximation still gives a clear indication to the inferiority of currently available portable capture technology.

3.3.4.1 Visual Acuity in AR

Relatively little attention has at yet been devoted to objectively measuring the effects upon the degradation that wearable devices impose upon perception of the real world itself. One study reported by Livingston et al. [85], see also Livingston [86] examined acuity measured using a Snellen chart at a range of 20 feet. Across a range of see-through

²see the following link for more detail <http://clarkvision.com/imagetail/eye-resolution.html>

displays it was found that the user suffered a reduction in visual acuity and contrast sensitivity by up to a half simply owing to the effects of viewing the world through the optics of the head-mounted displays.

3.3.5 Luminous Perceptibility

A persons luminous perceptibility or sensitivity to contrast is defined by the ability to discern differences in intensity (luminance) values across an image. The luminance of objects in the visual field of a HMD is especially critical when using optical see-through HMDs. These generally have lower luminance values than their WoW equivalents due to being subject to external light sources. In a task where brightness of objects needs to be satisfactorily within the limits of human contrast sensitivity such as well lit indoor or outdoor use it will often be more appropriate to use a WoW HMD so that contrast can be controlled by the designer. Determining this factor for a HMD can inform an understanding of how perception of the world is degraded. This measure of contrast is a discernible factor of visual acuity, a high contrast target will be easier to resolve than a low contrast target. To get an intuitive feel for how contrast sensitivity relates to acuity, consider the Campbell-Robson contrast sensitivity chart [87] reproduced in figure 3.5 from a mathematical description of the image. Looking horizontally from left to right, we see sinusoids of exponentially increasing spatial frequency, whereas from top to bottom we see a similarly logarithmic modulation of contrast, from 100% at the bottom to approximately 0.5% at the top . The inverted-U shape made by the sinusoids against a background of grey is an estimate therefore of the readers own contrast sensitivity function (CSF) subject to the degradation inflicted on the image by the readers own display equipment (if viewing digitally) or printer.

Perception of contrast reaches its optimum level at luminances of around 100fL^3 where contrast ratios of 0.02 can be perceived. The effect on human performance is a case of finding the effective contrast ratio in varying light conditions. Contrast perception is variant on the surrounding ambient light compared with the luminous power and contrast ratio of the device. This is particularly important in see-through HMDs. If

³foot lambert (fL) refers to a unit of luminance or photometric brightness, equal to the luminance of a surface emitting a luminous flux of one lumen per square foot, the luminance of a perfectly reflecting surface receiving an illumination of a one foot-candle.

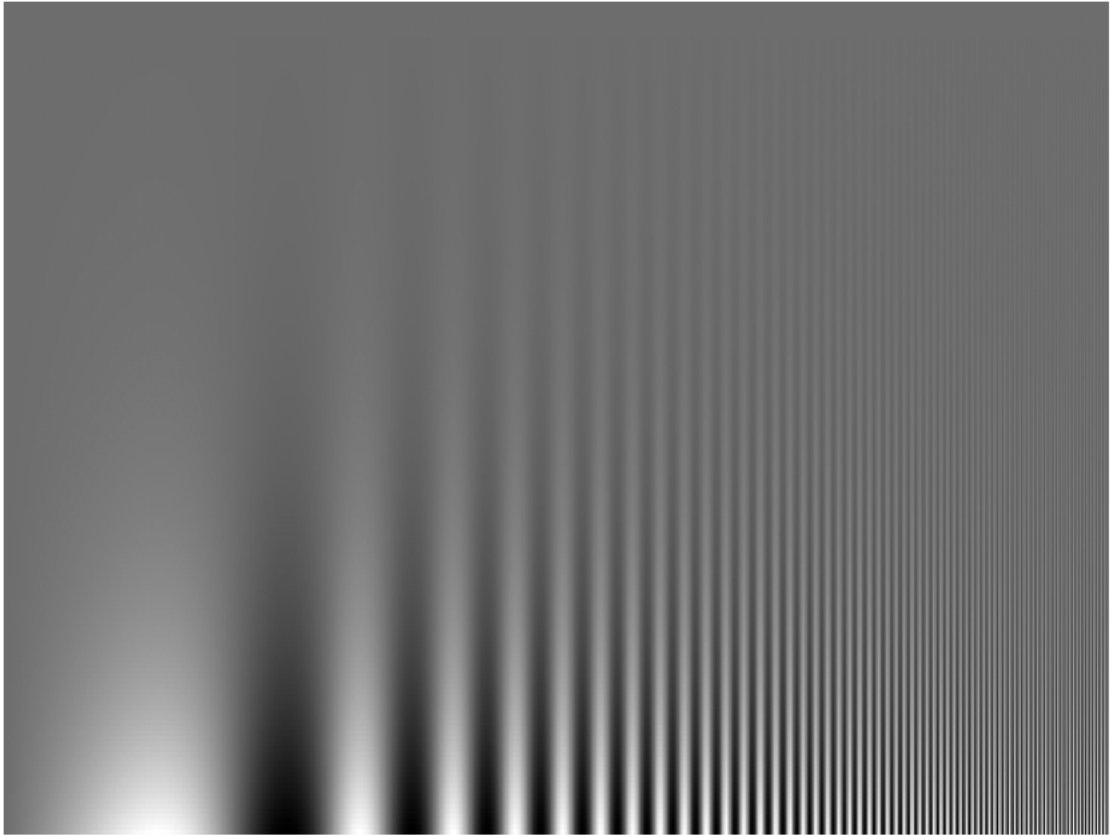


FIGURE 3.5: Contrast Sensitivity Chart. Reproduced from Campbell and Robson [87].

the light output from the device is not significantly higher than the surrounding light conditions any resulting images will appear ghosted or semitransparent.

Since both the real and virtual environments are available in digital form in WoW systems, very compelling MR environments can be created using chroma-keying technology. This allows for much more freedom across the RV continuum. Ghosting in optical see-through systems makes it very difficult to make virtual objects obscure real world objects. This is due to the fact that optical combiners allow light from both virtual and real sources [80]. The solution to this problem is to shut out light from the real world. However, this is very difficult to design as the projected object needs to be focused at two points in the optical path, at the users eye and at the point of the hypothetical filter. Since occlusion is one of the strongest depth cues the illusion of an MR environment is reduced to an awkward and un-compelling symbiosis of the two environments.

3.4 Spatial Cognition

The effect technology and media employed in any AR system can have on performance is difficult to scale. Using memory as a means to measure performance is a good way of evaluating performance because it is both goal orientated and to a degree self-directed.

All media, be they traditional or modern interactive digital presentations use spatial arrangements to organise information. Take for example the standard computer interface. This is known as a direct manipulation interface, an idea introduced by Shneiderman [88] and is more commonly often referred to as WIMP (windows, icon, menu, pointer). The basic design principal of this type of interface is to present information through a series of measures that users navigate and process easily; overlapping windows suggesting depth and folders are opened to reveal arrays of icons. Even the common computer mouse uses spatial motor interaction to translate motor movement over a 2D plane directly to a co-ordinate 2D visual interface. How information is represented spatially has been shown to directly facilitate cognition. Zhang and Norman [89] showed that subjects performance in solving the Tower of Hanoi⁴ varied dramatically based on the spatial arrangement of problem pieces. Kirsh [90] asserted that humans are constantly, both consciously and sub consciously, organising and reorganising space in everyday life to enhance performance. In his work he categorised spatial problem solving into three epistemic actions:

1. Spatial arrangements that simplify choices.
2. Spatial arrangements that simplify perception.
3. Spatial dynamics that simplify internal computation.

The key point that this wealth of research suggests is that environment awareness and learning has been demonstrated to be directly related to its spatial interaction. AR is arguably the most spatial of all media, unlike other computer interfaces the user interacts with the system through full body based motion within a volumetric space. What this means is that AR is a truly immersive medium because a user's perception

⁴<http://www.cut-the-knot.org/recurrence/hanoi.shtml>

of information arrays is egocentrically immersed. Mobile wearable AR systems have the potential to provide continuous digital support in real space. Real environments that can be annotated or enhanced digitally could be powerful in improving human cognitive activities, such as attention, planning, decision making and procedural and semantic memory [91]. Research done by Roy Ruddle and Simon Lessles [75], showed that human performance is improved dramatically by movement in real space. Their research tested users' ability to find hidden objects within a synthetic environment that facilitated both the rotational and transitional component of body based movement. Interestingly it was found that performance was improved by the transitional component of movement but not considerably by the rotational component. Possibly because of their direct association with VR technology, Ruddle showed that previous research focuses on the rotational component of movement and also increased visual fidelity. Ruddle's research showed that neither was important in improving human performance. Although Ruddle's experiment did utilise a VR system with body based movement, unlike AR this is not inherent to the technology. Little is known with regards to the effects of full body based movement with regards to the properties of AR.

3.4.1 2D vs 3D Interaction

There has been a great deal of prior work comparing the general effectiveness of 2D and 3D interfaces on spatial memory. An experiment by Tavanti and Lind [92], reported at InfoVis 2001, provides the most compelling result in favour of 3D - their participants recalled the location of letters of the alphabet more effectively when using a 3D interface than when using a 2D one. However, a later study by Cockburn [93] the results of which strongly suggested that the effectiveness of spatial memory is unaffected by the presence or absence of three-dimensional perspective effects in monocular static displays. In an immersive environment the user can move within the environment on any axis. Prior research has been approached within a limited domain, namely most research only addresses how the presence versus absence of perspective effects introduced by 3D interaction affects user performance.

Evaluating how the introduction of perspective effects in 3D environments affects spatial recognition has yielded divergent results leading to confusion over the efficacy of 2D vs 3D. The over-arching research finding has been that 3D displays are most beneficial

where the primary task specifically involves understanding terrain as part of decision making, for example judging lines of sight (St John et al. [94]; see also Lehtikoinen and Suomela [95] for similar findings). However, for tasks that require understanding absolute positions of units (such as aircraft in an ATC task) performance is better with 2D displays, possibly because 3D displays require observers to update their knowledge of positions along more axes (X, Y and Z; [96]) and 3D displays lack clear ways of imposing a veridical scales. Therefore the choice to use 3D virtual representations of objects over 2D should have some motivation other than simply enhancing spatial memory. In the scope of AR there are benefits to using 3D other than spatial memory namely, an AR system built around a 3D vector drawing engine is preferable to more commonly implemented 2D virtual object overlay methods since they replicate the real world experience more accurately and naturally resulting in a more immersive experience for users. Most importantly by representing virtual information in 3D human performance in the mixed reality medium will be more equally comparable to normal real world interaction.

3.4.2 Memory - The Use of Recall as a Measure of Visual Perception and Search

Recall has been previously used a metric of performance in synthetic environments (SEs), both to measure the impacts of variations in simulation fidelity and as proposed proxy metric for a sense of immersion (e.g., Mania et al. [97], Mania et al. [98], Dinh et al. [99]). The logic behind this is if there are deficits (as compared with reality) in how a simulation depicts spatial information within a scene, then this will lead to typically poorer (or at least different) mental representations of that scene that need to be accessed to perform a recall task. Furthermore, for many everyday tasks, there exist a number of studies that suggest an overwhelming impact of task upon self-directed eye movements (that is, as distinct from eye movements driven by stimulus properties such as contrast, chromatic salience, and motion [100]). Eye movements tend to be directed only to task-relevant stimuli and objects, and in terms of time-course objects only become task-relevant at the time during task performance they are required, equating to a just in time strategy for information pick-up, e.g., if we are making a cup of tea, the cup only becomes relevant after the boiling water has been poured into the teapot, indeed, during which time the eyes are fixated on the pouring spout of the kettle [101].

It has been shown that people construct a model of the space that they are searching. Moray et al. [102] states:

The observer constructs statistical models of the spatial and temporal properties of his environment. The observer uses the models both to govern the decisions he makes about the data obtained when he makes an observation, and also to decide when and where to make observations.

The notion of what constitutes a likely area changes with experience and illustrates that search is likely to be informed by a set of hypotheses concerning what one is likely to find and where one is likely to find it. The literature that deals with finding items amongst a clutter of objects is that which relates to visual search tasks. Generally speaking the findings are that accuracy falls as a function of the total amount of information displayed as studies show in Teichner and Krebs [103].

Studies conducted by Bacon and Egeth [104] show that where an item has a distinctive feature relative to distracter objects in the environment, the duration of visual search seems more strongly related to the distinctiveness of an item rather than the number of items in the search array. Therefore the ability of an item to 'pop out' during pre-attentive search is of critical importance and in some circumstances it can become easier to detect these items among distracters if the number of distracters increases. It should be noted that this factor in pre-attentive detection is only observed where there is a gross difference between target objects and distracter objects. Targets should be a singleton having a distinctive feature eg. being blue in an environment where distracters were red or green. Attributes such as colour also contribute to more time-intensive serial search by way of conjunction. Conjunction search occurs when multiple attributes are searched for ie. blue bottles marked with an X (as employed in experiment 3, chapter 7 of this research). In this type of search the number of targets displayed sharing relevant attributes is related to the time taken to find them [105]. There is however an important attribute to acknowledge regarding symbol search as observed by Treisman and Sato [106], that being that not all attributes are equal. Particularly salient attributes such as flicker, colour and size lead to quicker search time even where multiple attributes need to be searched amongst [107]. As such these salient attributes should be carefully considered at the design stage for any search based task so not to cause undue distraction.

The notion of conjunction is also observed in regards to search strategy. As the environment becomes more complex, studies in search strategy imply that recall is focused around the location of objects rather than their identity (see Beck et al. [108]). Parmenier et al. [109] asked participants to recall sequences of spatial locations and showed that path crossing and path length affected recall and response time. These studies imply a relationship between object location and recall, which suggests that manipulating location can have a bearing on the persons ability to recall an item (even when the objects identity is unchanged). A further limit implied in search of targets based task design is the limitations of human short term memory. This is commonly understood to be between five to nine items depending on the individual (see Miller [110]), also requiring users to memorise a large symbol set (ie. wide variety of items also leads to a degradation in performance [111]). The studies presented in this thesis require that users perform search tasks in which objects are presented in different locations. From this review it could be expected therefore that objects in certain locations would be more likely to be recalled than others. This raises a key question for the research: how is this relationship effected by technological media such as augmented reality?

One such effect could be observed due to the fidelity restrictions in currently available AR implementations. Thus, it is implied that generally speaking VR stimuli will be clearly distinguishable from the real world. This poses some interesting questions in regards to human performance in AR since the representation of virtual objects in the real world may cause indirect pre-attentive search with regards to AR vs real stimuli. As a result search performance in AR in regards to virtual objects may lead to improved performance over single modal mediums of target object representation such as real life and virtual reality.

3.5 Discussion

The sheer number of augmented reality systems detailed throughout this chapter present a paradigm of possibilities relating to the effectiveness of AR technology. Although it may never be possible to determine whether a users visual perception is fooled by an augmented reality system or is merely suspending disbelief, the studies examined have shown how task based approaches are helpful in determining the effectiveness of an AR environment. Proving how effective an AR system is in its presentation using low level

perceptual tasks is key to understanding how this technology affects visual perception and human performance [85]. Considerations for AR development as it relates to visual perception and thus human performance are summarised by the following:

1. Constraints imposed by display technology such as depth cueing, limitations to field of view, focus of attention and the ability to resolve visual imagery.
2. Direct aberrations to the visual field caused by tracking technology such as errors effecting orientation and/or pose estimation.
3. Physical load and human behavioural affliction imposed by AR system apparatus and design such as restrained head movement, binocular rivalry and/or limited depth of focus and causality of visual fatigue from prolonged eye fixation and/or displaced centre of gravity.

The limitations imposed by any configuration of AR system design are likely to have an effect on task performance. When compared to wholly real scenarios it is fair to summarise that AR will most likely have a negative effect on performance mainly because of the constraints imposed on visual perception. The findings in this chapter therefore imply that in any study on human performance, visual perception should be standardised as much as possible over the RV domain. Thus, in a comparison between real, AR and VR domains the world should be interpreted via the same display technology. The effect of linearising visual perception across various RV conditions will be absolutely crucial in determining how behavioural aberrations in rotational and translational movement inherent in AR technology affect task based human performance.

While much of the research on visual search makes use of eye-tracking to study search strategies, this review also indicates that recall can provide a viable outcome measure of search. One can assume that the ability to recall items that were found during a search task is, in some way, related to the search strategy and is therefore an effective measure of performance. While the recall of items can be affected by a host of other factors beyond search strategy, e.g., relationships between the items and the searchers mental model, conspicuity of items against a background, distraction and workload, it is assumed for the purposes of this thesis, that recall provides a usable surrogate of search activity.

The research done in this section suggests that in comparison to video based HMDs affordable visual see-through HMDs have poor contrast and luminance capabilities. This could impose severe compromises in terms of human performance due to the inevitable rivalry that would be created between perfectly realised real world imagery and possibly grainy, poorly focused virtually augmented information. Also, since a users depth and focus cannot be easily regulated or controlled, cognitive awareness of virtual information in the real environment cannot be guaranteed. This disparity between the virtual and real environment is highly undesirable since a measure in human performance cannot be substantiated where there are variables that have effects that cannot be confirmed or accurately regulated. In a case where the user is switching focus between virtual and real imagery the illusion of merged or mixed reality is precluded and thus a humans perspective on the world is not so much changed but simply split between information entities. To keep the effects to visual perception consistent across all RV conditions and in order to minimise disparity between RV environments the design choice implemented in this research comprises a monitor based display device coupled with a web camera to capture the environment.

Research shows that the majority of work in the field of AR has been achieved through the utilisation of fiducial markers placed in the real world that can be tracked using vision processing software techniques. This is possibly due to the affordability of the implementation and also the robust and accurate nature of the technology. While six DOF magnetic tracking offers the designer the opportunity to create dynamic feature rich AR environments, it also imposes drawbacks such as added weight, cost, reliability and imposes restrictions on freedom of movement. Another choice would be marker-less affine region tracking implementations that offer flexibility and usability in unprepared environment conditions. Despite how compelling this may be in concept marker-less approaches are currently less well documented, generally more primitive in terms of user performance and are susceptible to error in certain environments. For the purposes of the work done herein research suggests that visual fiducials is currently and certainly for the foreseeable future the most conventional and favoured implementation for AR systems. Therefore results pertaining to affects in human performance utilising a fiducial marker approach will be most useful in the present research community.

This thesis will further investigate the effects of AR on human performance through the implementation of a series of task based trials that aim to discover how the components of

movement inherent in AR influence performance when compared to real world scenarios. The experiment design will require users to search a prepared environment containing virtual and/or real static objects. In doing so it is hoped that something will be learned about the effect advanced AR implementations can be expected to have on human performance.

Chapter 4

Augmented Reality System Development

4.1 Introduction

The development and realisation of an augmented reality system requires that the designer make many key choices regarding system implementation, all of which may in some way introduce bias on human performance. Choices in both software and hardware implementations must be carefully considered before development commences. Human factors discussed earlier, relating to approaches in software and hardware are key in justifying the chosen approach. Key areas for choice of design include, tracking method, HMD type, and software implementation. Other constraints such as hardware/software cost and integration also have to be taken into account.

4.2 Hardware

Research of literature in earlier chapters suggest that the HMD hardware is primarily responsible for many factors pertaining to visual perception within an AR environment. At this stage in development therefore, the most important factor affecting the design proposition was the choice in regards to the type of HMD hardware. For all intensive purposes a system based on a WoW HMD technology was favoured over a see-through system. A visual see-through HMD approach was deemed less appropriate not only

because of the limitations imposed by affordable hardware but also the restrictions pertaining to the control of the user environment. Although a video based method does dramatically degrade the fidelity of the real environment, this environment can be controlled and the users viewing perspective can be focused more easily. Also since the fidelity of the real world view is limited by the world capture device, equally resolved virtual data can be augmented and easily matched to the resolution of the real environment as it is captured. This creates a more compelling and plausible mixed reality experience because the merging of mediums is indistinguishable. Note that this implies, there is a trade-off between degradation and fidelity between the realism of the real environment and effectiveness of the AR environment. Although the research suggests that simply a loss of fidelity between conditions is unlikely to affect human performance. Any inconsistency between HMDs could introduce a bias to visual perception. To avoid introducing a bias on any given experiment the same optics should be used between each RV condition. This ensures that the effect of AR on human performance is due to AR properties and not visual disparity.

4.3 Tracking

In order to annotate virtual information, systems that use pattern recognition, such as the ARToolKit based system developed for this research require that an environment be prepared. Systems have been developed that successfully track geometric features in the real world, such as picture frames, door handles for indoor use or building features coupled with GPS positioning for outdoor use (see studies by Ferrari et al. [66] Neumann et al. [112]). What all recognition tracking systems have in common is that they estimate pose using objects placed in the real world. The tracking system is principally the core of an AR system. One possible and unexplored approach discussed here looks at the viability of using real objects as markers on which to annotate virtual stimuli and thus create an AR experience. Discussions and talks involving an inventory system for forensic case work arrived as a novel use for an object recognition system that could categorise real world objects and then also allow for annotation using AR. In real world case work, large quantities of evidence can be collected from a crime scene. A system trained to recognise and annotate real world objects with additional relevant information could be utilised to aid a CSI in linking evidence to a narrative of the crime (see chapter 1). Such

a system could be employed to enhance evidence on an evidence review table or be used in situ with objects at a scene of crime.

4.4 Software - Windows Presentation Foundation (WPF)

The literature presented here introduces the Windows Presentation Foundation (WPF) as a favourable software platform for creating a readily adoptable 3D based augmented reality system and puts forward a concept that allows for the whole AR system to be developed under a multitude of separate and dynamic entities.

Responsiveness to its environment and its inherent interactivity has generated a development trend for the world wide web to become a multipurpose application platform. The requirement for the internet platform to be as flexible and dynamic as a client based application has become increasingly more apparent in recent years with significant gains in the performance of hardware architecture and internet speeds that are rivalling home computer networks. This need to control the presentation of linked media, receive input, display multiple languages, obtain information from the runtime environment, acquire dynamic content and invoke client-side logic is difficult to satisfy using the conventional programming model. In modern computer systems the web experience is placed in an environment that is mostly detached from the often very capable hardware that is providing it. This could be problematic for a graphically demanding system that required shared user input over an internet platform such as an AR interface. The Windows Presentation Foundation (WPF) is built into the core framework of the Windows Vistas graphics platform. Anything that can be described by its programming model XAML can be presented by Windows Vista and visa versa [113]. With XAML a new paradigm for programming is presented to the developer because XAML is a description of the user interface, not an abstraction. XAML takes what has been done with other tag based languages such as HTML and XML and applies it to the windows interface, in doing so it is possible to represent any windows object as an XAML tag allowing for very complex and compelling user interfaces to be created.

The emphasis in Vista is for domain specific language environments so that the gap between UI design and programming logic development may be bridged. The application of such technology as regards to AR is compelling because the 3D AR environment base

can be developed wholly using style descriptors and 3D vector mesh co-ordinate data in XAML. Once the UI interface is developed a programmer can hook up the required logic for the tracking and registration system in C# or VB.net. The importance of this distinction is that in the conventional software development model programming languages such as C++, C#, Java and Visual Basic co-exist with script based languages such as HTML, XML, VRML and Java Script in an uneasy symbiosis. Declarative script based languages known more commonly as mark-up languages are tailored towards creating a good user experience that is highly textured and adapts to different screen sizes and environments. However, they are hopelessly inept at providing user interactivity in a non-trivial way. Inversely programming languages can be tailored to do anything the computer hardware is capable of but rarely provide visually enticing interfaces because of the added programming load it puts on developers. WPF provides a strategic graphics subsystem that provides a unified approach to the user interface and because it utilises Direct 3D for vector-based rendering, powerful solutions for building immersive applications can be realised by unleashing the full potential of the graphics hardware present in almost all modern computers [114]. Through the unification of browser-based experiences, forms-based applications, graphics and video, audio and documents and the way in which designers and developers experience these holds great potential for compelling and sophisticated AR systems development.

4.5 Initial System Design Concept and Development

One of the key aspects of accurate registration in an augmented reality system lies in location-aware computing [53, 115]. Much work has been done with systems that adopt a hybrid of tracking technologies either simultaneously or in alteration, depending upon the environment [116]. The system based on visual fiducials used for this research and described in earlier (see 2, 2.5) is the most popular approach to location-aware computing. The concept of the system developed here uses real world 3D objects as markers for location-aware computing, and appends the object with virtual information via a head mounted display or HMD. In the context of a crime scene this would allow for the investigator to enhance their understanding of the evidence with very little or no intervention thus, creating a ubiquitous computing environment.

The first step to take when trying to analyse an image is to make the object for recognition as simple and easily definable as possible. Such is done with the fiducial marker system developed for this research and other such systems such as Wagner et al. [117] that utilise the AR toolkit designed by the Hitlab at Washington University. For this system we are interested solely in classifying the basic shape of a real world object, thus a simple outline of the shape in question will suffice. With this realisation our first step is to find a way of removing as much as possible, all noise (information we are not interested in) in the image and find the outline of an objects basic shape.

4.5.1 Edge Detection

One of the most commonly used pre-processing operations in image analysis is edge detection [118]. We define an edge as being a discontinuity in grey level values, in other words an edge is the boundary between an object and the background [119]. The shape of an image edge is determined by the geometrical and optical properties of the object, luminous conditions, and the noise level. Since edge detection is the first step in image processing it is important understand its fundamentals and know the differences between edge detection algorithms. Of the many types of edge detectors available such as; zero crossing [120], Laplacian of Gaussian, Gaussian and Coloured [121]; gradient edge detectors have been found to be the most effective and most widely used in image processing because of their simplicity and computational inexpensiveness [122]. For these reasons some gradient based edge detectors were chosen namely Kirsch, Prewitt and Sobel and implemented.

4.5.1.1 Testing and Results

As can be seen from these results (see figure 4.1), edges are most vividly defined when they are being identified by the Kirsch Operators. In fact the Sobel and Prewitt operators would be problematic as a first step towards object recognition because in areas where lighting is poor the gradient magnitude on an edge is not significant enough for these algorithms to recognise and define a pixel as an edge pixel and hence the full edge of the object is not recognised.



FIGURE 4.1: Comparison of edge detection operators

4.5.2 Automatic Machine Recognition

Automatic machine recognition, description, classification and grouping of patterns are important problems in computer vision. A pattern is defined by Watanabe [123] as:

"the opposite of chaos; it is an entity, vaguely defined, that could be given a name."
Watanabe [123]

Therefore a pattern is for example a fingerprint image, a human face, a handwritten cursive character, a bottle or a speech signal sound wave. Given a pattern, its classification should comprise of one of the following [124]:

1. Supervised classification in which the input pattern is identified as a member of a pre-defined class.
2. Unsupervised classification where a pattern is assigned to a hitherto unknown class. Here the recognition problem is being posed as a classification task. Usually these systems are learn based on the similarity of patterns.

And furthermore given the classification method a pattern recognition system will essentially comprises three key aspects [124]:

1. Data acquisition and pre-processing.

2. Representation of data.
3. Decision making.

Pattern matching can be achieved by employing any of a number of algorithms, such as;

1. Template Matching - The pattern to be recognised is matched against a stored template while taking into account factors such as scale and rotation. This method is fairly rigid and computationally expensive to implement. It also performs poorly when there are large intraclass variations among the patterns and where patterns are distorted due to noise [125].
2. Statistical Matching - Each pattern is represented in terms of d features and is viewed as a point in a $d - dimensional$ space. Given a set of training patterns from each feature class the object is to establish decision boundaries which separate patterns from each class.
3. Syntactic Matching - Where it is necessary to recognise complex patterns, it is more appropriate to use a hierarchical approach where a pattern is viewed as being composed of simple sub-patterns which themselves are comprised of even simpler patterns and so on. The implementation of a syntactic approach is difficult as in many instances it can yield a combinatorial explosion of possibilities to explore before successful classification causing a demand for very large training sets and computational effort [126].
4. Neural Networks - see section 4.5.3 below.

The challenge to reproduce human like response and aptitude for mundane tasks such as pattern matching in an artificial system is now well recognised. Humans are able to distinguish large characters, small characters, hand written, machine printed or rotated characters easily from a very early age. And these characters may be written amongst a cluttered background, on crumpled paper or even be partially occluded. It is still not understood how humans recognise patterns, however it is well documented that the human brain is modelled most closely by neural networks. So despite their distinct similarity to statistical pattern matching approaches neural networks offer several advantages unavailable to other methods such as; unified approaches for feature extraction

and classification of data and flexible procedures for finding good, non linear solutions. Also with the advent of studies that are beginning to prove that the way in which the human brain accesses stored memory can be described by chaos theory and catastrophe theory, it is possible that ultimately a system that marries chaos theory with neural networks could create a very accurate pattern matching system model. So whilst not the most popularly adopted method for pattern matching tasks the development of neural networking methods is offers a compelling challenge to developers.

4.5.3 Neural Networks

The first step towards true parallel computing systems is neural networks which are made up of many simple processing nodes (usually in the millions) called perceptrons. These perceptrons are each interconnected to every other perceptron via weighted links [127]. The value of these weighted links affects the behaviour and input/output response of the network. The main advantage of a neural network system is that it possesses the ability to model complex non-linear input/output relationships and adapt its behaviour to the data using sequential training procedures. The most popular neural networks in use for pattern recognition tasks are the feed-forward back-propagation network (BPN) and the Kohonen Self Organising Map (SOM) that offer supervised and unsupervised classification respectively. In spite of attacks from statistics buffs claiming that neural networks are simply statistics for amateurs because in many ways they are implicitly equivalent to classical statistical methods of pattern recognition and they conceal statistics from the user [128], neural networks have received renewed and ever increasing interest in the last few years. This is due to their low dependence on domain specific knowledge. Another reason for this renewed interest has come about from the creation of cellular automata machines which are able to exploit the unique parallel nature of neural networks [129]. In terms of image processing this means that the computational overhead governing a 640x480 is no different to that of a smaller 320x288 image [130].

The basic concept of a neural net is made up of layers, namely an input and output layer, but a set of one or more hidden layers may also be used. Each node from a layer will connect to each and every node on the next layer (see figure 4.2) [131].

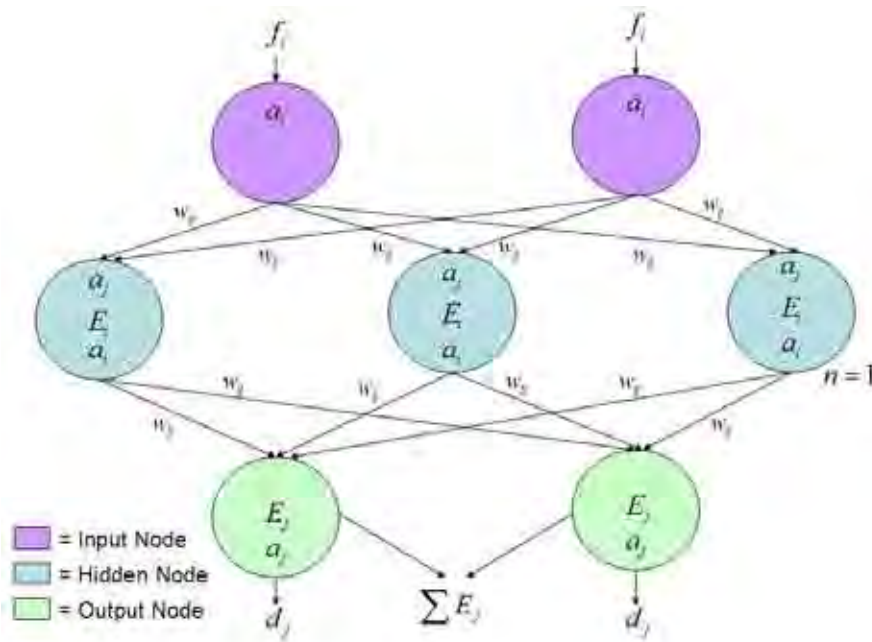


FIGURE 4.2: Neural Network Node and Layer Structure

The network is first initialised with random link values between -1 and 1. At this point the network has no useful function so the next stage is to train the network, which is the process of feeding the network input data with known output data.

4.5.3.1 BPN Applied to Pattern Recognition

The first stage was to specify the number of nodes needed for each layer of the network. We start with a set of training images and then can specify the number of input layer nodes to be equal to the number of pixels in the given image; the number of output layer nodes is equal to the number of images in the training data set and a good approximation for the number of hidden nodes was found to be $(\text{Input Nodes} + \text{Output Nodes})/2$ contrary to Kolmogorov's theory that specifies $\text{Hidden Nodes} = 1 + 2(\text{Input Nodes})$ [132]. A value for L must also be chosen; a low value will increase training time but increases the performance of the network. A series of trials on test data showed that a value 0.2 was found to allow the network perform with satisfactory performance. The process of presenting a set of training data to the network is called an epoch, the network is considered trained when the total error for each epoch is below a specified threshold value, say 0.1. A well trained network can be achieved by applying the following procedure. The first stage is complete when the network correctly recognises all

of the training data correctly within one epoch. In the second stage we follow the same procedure as step one but present the data set in a random order, then finally the last step requires us to process as many epochs as it takes to get the total error below the specified threshold value [133].

4.6 Neural Network AR System Analysis

Using the systems developed to enhance the important features of the image (edge detector) and differentiate that image from another (neural network) it is possible to create an AR environment. A group of images are prepared using an edge detection operator, these images are then loaded into the neural network which is then trained. The next step is to re-present these images to the neural network. A webcam is used to capture these images from the real world, they are processed by the edge detection operator and then presented to the neural network and matched to the correct item. Based on the matched pattern an image is sent to the HMD that presents the user with a description of what they are looking at (see appendix D).

The program (see figure 4.3) allows the system operator to load in upto six images, train them into the network then load in test images to see if the correct output node is selected. It is basically a testing platform that proves the neural network implementation has been successful.

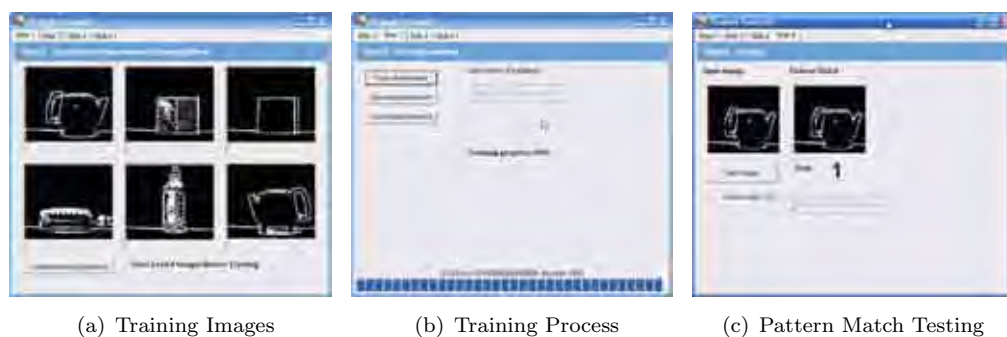


FIGURE 4.3: Neural Network Pattern Recognition System

Six prepared images were trained into the neural network. The real objects are then placed in front of the user, at the click of a button the user sees the object augmented with object specific information simply overlaid on the input image. The tests were quite successful however the testing environment had to be quite rigid in order for the



FIGURE 4.4: Upon successful recognition the user is presented with contextual information specific to the object

pattern matching to be successful. If an object was placed in an orientation and position that differed in pose from its trained counterpart then often the system would fail to make a correct match. It was however possible to mitigate these issues using more granular symmetrical objects such as bottles, paper cups, dice etc and further tests with such objects gave promising results. Despite attempts to restrain parameters of use participant pilot studies (see figure 4.5) deduced that requirements of the hardware did not satisfy that of the user. Users reported discomfort with the system because generally there was a 1-2 second delay in registration and the registration process involved manual input from the user (ie. button clicks) and any errors in registration would often lead to a compromised result. This arguably led to a disbandment from the autonomous intent of the system. The problems with this system are inherent in the way the neural network analysis captured the array of pixels, any variation caused by situating the object with a large enough offset or with a differing orientation appears to be a different arrangement on the input layer of the neural network. Further development of the pattern matching system could help improve the system flexibility and performance. Such improvements would be achieved by incorporating feature extraction algorithms to recognise features of an image before analysing its pixel arrangement. Even with these improvements there would still almost definitely continue to be a pattern recognition bottleneck resulting in compromised registration in a dynamic moving environment.



FIGURE 4.5: Pilot User Trial - User looks at real world object and is presented with object relevant information.

If it was felt that although further improvements to the system could overcome the performance issues there was still a risk that the need for a user to control the technology would undermine the the research goals of this work. The idea that augmented reality could be used sufficiently to aid training platforms as an alternative to wholly real life platforms is completely undermined if the AR technology itself does not lend itself well to natural interaction in a real environment. What the development and testing of the neural network based AR system suggests is that even with a great deal of improvement the users need to control the technology such that registration is not compromised. This as such is a contradiction in terms of what AR as a concept and extension of ubiquitous computing environments is supposed to achieve. This deduction also lends itself to other approaches to AR that use registration methods that are sensitive to the environment such as affine region tracking, magnetic and IR based systems. Possibly the most robust and well learned approach to tackling registration in AR involves adopting a fiducial marker based approach utilising the ARToolKit. This software SDK written in C++ is generally the de-facto standard employed by most implementations of AR. A system built on the back of this framework offers a good benchmark as to the performance of AR in its most widely adopted form.

4.7 WPF Software Design

Development of a rich 3D vector based client environment using the WPF framework was achieved utilising Direct 3D technology. Direct 3D is a graphics engine that is integrated directly into WPF. This gives direct access to hardware allowing the developer to take full advantage of the Direct 3D API within the WPF framework. The means by which this power has been brought to the developer is really quite innovative. The architecture of the WPF subsystem is inherently the graphics engine behind the Windows Vista platform. Ultimately responsible for setting pixels on screen is `Milcore.dll` which processes the required 2D and 3D imagery using tessellation and texture management so that it may be comprehended by Direct 3D. `PresentationCore.dll` provides APIs for communicating with `Milcore.dll`. This process called managed graphics is where animation is rendered separate to the main User Interface (UI) thread. Whilst the UI thread processes objects that the user interacts with directly such as visuals, buttons and windows, a composition engine runs in the background managing the graphics; performing tasks such as refreshing the display frame and handling animations. The advantage of managing graphics in this way is that the protocol on which they communicate can be message based and therefore the composition thread can be run on a remote client machine with no noticeable performance loss. This means that rather than having to pipeline alpha blended graphics pixel by pixel to the remote client such as is the case with for example Java 3D and GDI+, instead using a fraction of the bandwidth all that is sent across is the developer intent thus allowing for complex graphics tasks to be performed on the remote client.

The most popular implementation of camera tracking for pose estimation is built on the framework SDK developed by the HIT lab at the University of Washington. First the image is converted to a binary image (see section 4.5.1) then when a pattern is recognised by way of syntactic pattern recognition techniques (see section 4.5.2) a matrix transformation is returned which defines the detected markers position and orientation (6DOF) relative to the cameras calibrated co-ordinate system. The camera calibration data is then used to modify these results to compensate for the cameras unique properties. Due to errors in the calibration process and distortions in the camera the co-ordinate system is not the typical orthogonal co-ordinate system used in 3D graphics. When we get the transition matrix from the camera we get the marker relative to the cameras

co-ordinate system, in order to get the position of the camera relative to the marker this transformation needs to be inverted. ARToolkit uses a calibrated camera perspective that typically results in an off-axis projection matrix for OpenGL, defined as the image distortion function. Thus, rather than decomposing ARToolKits projection matrix into parameters we pass back to DirectShow and load the OpenGL projection matrix directly, in doing so no matter how poor the camera calibration is, virtual data pose will be correct because the calibration model has been reversed when rendering using the camera as the view frustum. More modern extensions and re-written frameworks of the ARToolkit that allow for more efficient operation and more accurate estimation of fiducial markers for camera pose estimation include: ARToolkit plus [42], ARTag [134] and more recently Studierstube Tracker [135]. All of the aforementioned frameworks were evaluated. The chosen implementation is ARToolkit. Other toolkits such as ARToolkit plus is an extension of the commonly used ARToolkit whereas ARTag [136, 137] and Studierstube Tracker are completely re-written frameworks. Support in the development community for ARTag and Studierstube is still lacking when compared with the more popular ARToolkit, nevertheless support for ARtoolkit plus is also poor. The ARToolkit plus utilises more efficient algorithms for mobile computing use with devices such as mobile phones and documentation can be carried over from the ARToolkit. Utilising ARToolkit plus is beyond the scope of what is required for this research since mobile devices are not being utilised however, the system developed could be modified in the future to incorporate the additional efficiency offered by ARToolkit plus.

4.7.1 2D or 3D

A 2D implementation of augmenting virtual data can be achieved quite easily by simply scale drawing 2D virtual objects onto the view frustum of the real world input stream, however chapter 3, section 3.4.1 showed that a 3D implementation was preferable for AR. From a development perspective and for the majority of system designs the implementation of a 3D environment is far more taxing than 2D. Until now, creating a 3D environment has required that a very involved programming venture be undertaken and demanded a steep and often timely learning curve for even an experienced programmer new to 3D programming. Many 3D based AR systems to date have been developed using the 3D environment provided by the ARToolkit. This system is programmed natively

in C++ and uses DirectX to render 3D objects, however in order to make development easier 3D object descriptions can be provided in VRML using a DirectX parser. Despite this, the learning requirement for Direct3D is still required. The WPF environment provides the efficiency and robustness of DirectX 9ex edition (codenamed WGF 1.0) and allows for rapid development through use of its script based graphics module XAML pronounced zammel. This essentially eradicates the design process allowing for development focus on other logic tasks such as accurate tracking and registration.

4.8 3D Graphics Theory

3D objects are defined by meshes. A mesh is basically a representation of a surface. The mesh represents the surface through a system of points and lines. The points describe the high and low areas of the surface, and the lines connect the points to establish how you get from one point to the next. At a minimum, a surface is a flat plane. A flat plane needs three points to define it. Thus, the simplest surface that can be described in a mesh is a single triangle. It turns out that meshes can only be described with triangles. That is because a triangle is the simplest, most granular way to define a surface. Therefore a large complex surface cannot be accurately described by one triangle. Instead, it can be approximated by many smaller triangles. One could argue a rectangle would be a more efficient way to define a surface since less could be used to approximate the same surface, however this is not preferred because a rectangle is not as granular as a triangle, hence a rectangle can be described by two triangles but the same is not true for the opposite scenario.

A whole mesh is composed of:

- Mesh Positions - A mesh position is the location of a single point on a surface. The more dense the points are, the more accurately the mesh describes the surface.
- Triangle Indices - The mesh positions alone cannot describe the mesh triangles. A triangle index is a mesh position that defines one of the three points of a triangle in the mesh. After the mesh positions have been added, the indices of these positions that make up each triangle need to be defined, hence triangle indices.

- Triangle Normals - After defining positions and triangle indices, normals should be added to each position. A normal is used by WPF to know how the surface should be lit by a light source. A normal is a vector that is perpendicular to the surface of the triangle. The normal vector is computed as the "cross product" of two vectors that make up the side of the triangle.

4.8.1 3D Viewport

A 3D viewport specifies where to draw 3D objects in the active virtual buffer. The view frustum of the 3D viewport is defined by an interaction between field of view (FOV) and position of a camera placed in virtual space. To render and manipulate a 3D viewport scene using a WPF perspective camera three properties are manipulated:

- Field of View - The horizontal bounds of the cameras projection between the cameras position and the image plane.
- Position - The X,Y,Z co-ordinate vector location upon which the cameras projection is centred.
- LookDirection - The orientation vector, described by pitch, roll and yaw upon which the cameras projection is directed.

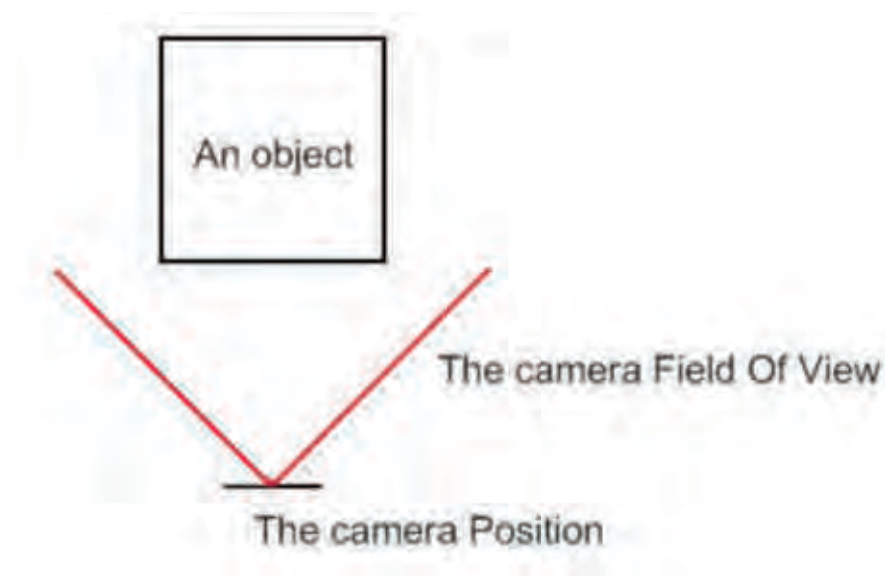


FIGURE 4.6: A Camera's FOV Relative to its Position.

Changing the position of the camera with respect to the view frustum affects the scaling ratios of geometry in the z (depth) domain and hence emulates a perceived perspective of depth. AR can therefore be achieved through manipulation of both position and orientation values with respect to a matrix transformation of a real world fiducial markers pose captured and estimated via tracking software. In order to augment the data more accurately in the users view frustum, the virtual cameras field of view should also be set such to mirror the FOV of the real world capture camera.

```
84 <PerspectiveCamera x:Name="camera"  
85     Position="141.7789546,-16.3768236,46.4545218"  
86     LookDirection="33.8546692379858,22.4674765098457,-14.6998457850988"  
87     UpDirection="0,0,1"  
88     NearPlaneDistance="5"  
89     FarPlaneDistance="600"  
90     FieldOfView="40"  
91 </PerspectiveCamera>
```

FIGURE 4.7: Sample Code in XAML, Description for a Perspective Camera.

4.8.2 Direct Show and WPF

Direct show is a multimedia API framework developed by Microsoft that provides developers with libraries and tools to perform various operations on media files or streams (it replaces the earlier VPW or video-for-windows technology). DirectShow is based on the Microsoft Windows Component Object Model (COM) framework which provides a common interface allowing for implementation across many programming languages. A filter based subset of libraries allow for media to be streamed, rendered and recorded from a capture device at the behest of the user.

WPF provides the functionality to render media streams into the 3D viewport as the background to the view frustum or as a dynamic texture to 3D geometry, so for instance a remote media stream such as a streaming video broadcast can easily be rendered on to the faces of simple or complex 3D geometry. Unfortunately however, no native functions are provided to render media streams from a local capture device such as a webcam. This is because low-level hooks to directshow composition have yet to be integrated into the direct show framework.

Since WPF does support internet based media streams and these streams can be setup to point to a local capture device a work around can be created where the WPF media

element is used to reference a live capture device set-up as a live internet stream on a local TCP port (such was the method adopted for the neural network based implementation). There are clear reservations to utilising a method that has to subvert captured media streams through a variety of protocols before it can be utilised and some of these reservations are because it places increased load on processing bandwidth and introduces lag on the video feed. The reason for thus far, precluding native access to directshow in WPF has yet to be explained and substantiated by microsoft, however it will inevitably be integrated in the .net framework in a future version. Another approach would be to generate a standard windows form control as a background thread and then display the media stream in a windows forms derived video window. The problem with this is that the video element then becomes protected within the thread created by windows forms, thus WPF cannot access this pixel data and hence 3D geometry cannot be drawn over the media stream via the WPF 3D viewport.

Fortunately the directshow library can be implemented into the WPF framework via the use of a C# wrapper. Since direct show provides a plug-in type architecture the COM object can be interjected with filters. This is how codecs such as DIVX are installed. The media element in WPF utilises a HTTP protocol handler to play local or remote media streams. What this means is that a custom protocol handler can be developed and in this way the WPF media element can simply render a media stream from a local capture device. There is however still a problem with this method. In order to track the fiducial marker the transpose of the camera matrix and the pixel data from each frame must be passed to the AR toolkit. Using a custom protocol handler allows efficient rendering of the capture source within the framework of WPF. However, access to the filter graph that Directshow creates is still protected.

4.9 AR System Composition

Currently the most effective and least computationally expensive way of accessing the directshow scene graph is to access the native image buffer. WPF uses the windows imaging component (WIC) to handle imaging in the user interface viewport. In order for WPF to reference images managed by WIC a handle is needed. This handle is located in the bitmap source class, modification to this class to lock the image pixels and get the native pixel buffer allows for a pointer to the image pixels. Access to the native

bitmap source buffer allows for pixel information to be passed to the ARToolKit for processing and thus, virtual data can be transposed in the 3D viewport with the correct pose and alignment. To summarise, directshow isamplegrabber callback provides a pointer to the capture device stream, this is then copied directly to the bitmapbuffer via the bitmap source class. This sample image is also passed to the ARToolKit tracker class which passes back the transform matrix and detected marker ID to WPF. The AR toolkit returns a camera matrices for the pose estimation of a real world marker in OpenGL format. WPF then sets the 3D viewport virtual camera projection matrix and draws the xaml geometry description. Both OpenGL and WPF use a right hand co-ordinate system. This is very helpful because the matrices returned from the tracker can be transposed directly into the matrix transform for the virtual camera. There is however, a shared airspace issue in WPF because the video layer resides within a windows form control WPF elements cannot be composed on top of it. The existence of this issue is due to the aforementioned lack of support for directshow in WPF. Fortunately there are workarounds available, using a BitmapBuffer hack paired with a DirectShow ISampleGrabber Callback. The ISampleGrabber Callback provides a pointer to the captured image then just copies that memory directly to the BitmapBuffer sample.

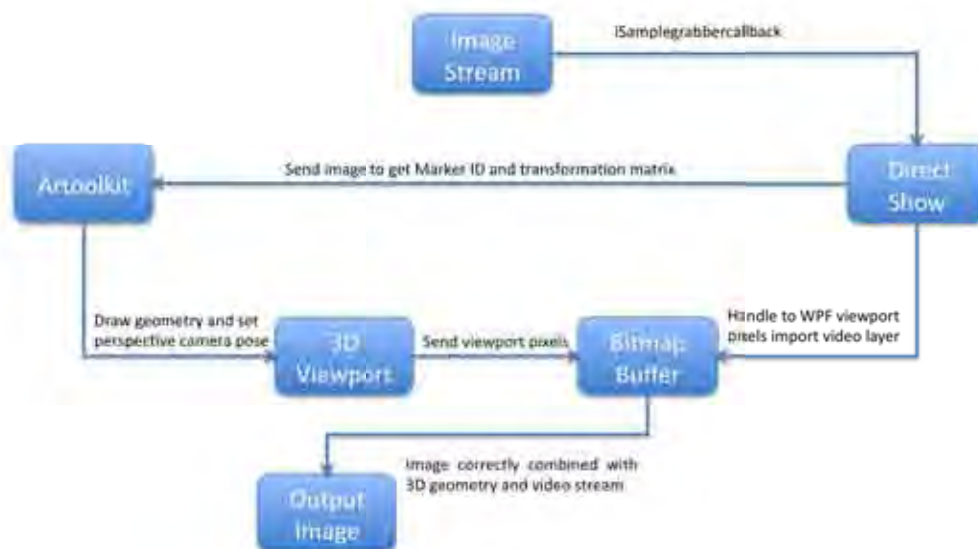


FIGURE 4.8: Architecture of AR System

4.10 Discussion

The applications for the neural network system developed here are vast, an application relevant to the forensic domain could be to aid research of various items of evidence from the evidence inventory. The user simply looks at a piece of evidence through the HMD device and is presented with relevant information such as case number, discovery date, owner information and is presented with graphical data that is augmented onto the item such as a finger print or blood stain. Another unrelated but equally profound application of this system could be to recognise items of food and then present the user with nutritional information. Imagine this system built into a simple camera phone (quite possible with the advent of new phones achieving pixel resolutions of upto 5 megapixels) the user sequentially inputs each item of food to be used in the meal they are to prepare, the system then computes all the nutritional information and outputs recommendations specific to the user (note this is especially useful if the user suffers from some form of anaphylaxis) eg. warning Charad's processed beans contain traces of macadamia nut.

Since the core of the neural network based AR system developed is based on image processing, more specifically image recognition it is therefore riddled with the problems inherent in all image processing systems. Thus the development required to make the system accurate and reliable enough to be used for the applications described earlier is a prodigious task. Continued development of this system required further work namely because of the problems inherent in the image processing domain and also the system developed here is focused towards an end product that works based on a static review of evidence in an AR environment, the system would have to be supplemented with further vision processing techniques in order to be utilised in a dynamic environment. A review of the research concluded that further development of this complex system was not concordant with the research goals of this work and therefore it was decided that it would be more appropriate to continue the research utilising a fiducial marker system.

Using the ARToolkit, incorporated into the WPF .NET framework a fiducial marker system was developed. What the composition of the WPF framework offers to AR development is quite pre-eminent and its virtues certainly in respects to its low bandwidth and scope for remote 3D functionality could be used to offer very powerful 3D context

aware environments that stream dynamic 3D content straight to the user space whenever the client technology enters AR domains. In terms of a crime scene application for AR this ability to stream active content between users could be helpful in building a narrative between scene investigators and data analysts.

Chapter 5

AR Performance Study - Frame of Reference

5.1 Introduction

Take two scenarios, one in which a catalog of evidence is viewed sequentially one after the other and a second scenario in which evidence is placed in a per-clockface fashion around the user's person. Logic would dictate that the second condition should enable the user to study and recall this information more accurately because the information can be appended to a spatial frame of reference in the real world. This study attempts to test this claim that the presentation of virtual information, if superimposed onto the real world, provides a reference frame that improves human performance. A study into frames of reference by Mou and McNamara [138] showed that users have a natural tendency to use an environment stabilised reference frame to recall the location of objects but that with practice a body stabilised reference frame can also be used. An advantage of using AR to represent virtual information is that it enables the user to view this information with a real world reference frame. A fair hypothesis for this experiment therefore;

AR enhances human performance by providing both a body stabilised and environment stabilised frame of reference.

To test this theory requires examination of two states:

Body Stabilised Frame of Reference (BSFoR) this is tested via facilitation of movement across a horizontal plane.

Environment Stabilised Frame of Reference (ESFoR) this is tested by controlling the availability of a real world background within the users view frustum.

Conversely it could be argued that a frame of reference acts to distort and detract from the observed stimuli, suggesting that human recall performance will be improved if objects are presented exclusively in a spatial arrangement (AR Off). There is a great deal of research that supports this argument in the context of 2D and 3D interfacing, where it has been shown that efficient use of graphical user interfaces is strongly dependent on human capabilities for spatial cognition [92]. One facet of spatial cognition is the ability to quickly and accurately recall and access the location of objects in a spatial arrangement [93]. Taking these findings into consideration the effect of a real world frame of reference has, poses an interesting and important question relating human performance to AR. Such being, that this movement associated property inherent to AR technology may be instrumental in aiding task based human performance.

Experiment 1A The examination of the two states (BSFoR and ESFoR) required that a total of four experimental conditions be devised. Firstly two conditions are tested (called the movement condition) where a body stabilised frame of reference is always provided but the environment stabilised frame of reference can be turned on and off. Then secondly a further two conditions are tested (called the static condition) where the environment stabilised frame of reference could be turned on and off but no body stabilised reference frame is provided.

Experiment 1B - Blocking Condition It was felt that the chunking of the results data into three discreet sets 'primary, mid-range, recency' made an assumption that these values belonged together. Thus a condition was devised to mitigate the serial position effect (SPE) and verify the validity of the findings made in experiment 1A.

This chapter aims to satisfy the following research statement:

The review of literature implies that movement is a principle factor affecting human performance. How does this apply in AR?

5.2 Participants

In total, twelve subjects each took part in fifteen trials of each condition. All participants were university graduates aged between 21 and 30 years of age with an average age of 25. Of the twelve participants 9 were male, 3 were female.

5.3 Hardware Design

The nature of this experiment requires that the user be fully immersed in the AR environment to enhance the disparity between each condition and also to ensure that the user is subjected exclusively to the visual stimuli intended by the designer. A binocular video based HMD is therefore an obvious first choice to meet this criteria.

HMD *Daeyang i-Visor PC SVGA DH4400VP*

The chosen HMD (refer to table 2.1, chapter 2) utilises 100% stereo overlap and therefore is suitable for monocular input, hence a single camera mounted in the central field of view can be deemed suitable for this experiment (see figure 5.1).



FIGURE 5.1: Daeyang i-Visor PC SVGA-3D Pro DH4400VP Head Mounted Display

Capture Device *Creative Labs Live! Ultra VF0070*

The Creative Labs VF0070 (see figure 5.2) is a compact camera designed for notebooks and therefore boasts a suitable form factor and low weight. The camera has a resolution of 640×480 utilising a CCD sensor and adjustable lens. Many cameras considered provide a field of view of 50° or less, the wide angle lens on the VF0070 offers a modest 76° .



FIGURE 5.2: Creative Labs Live! Ultra Webcam

Computer Hardware *Dell XPS M170 Portable PC* The Dell XPS is a portable computer with a 2Ghz Intel Pentium processor which boasts a powerful onboard graphics display adapter (NVIDIA GeForce 6800 Ultra) allowing for smooth performance of DirectX 3D media.

Static Condition - A second desktop PC was used with a wireless USB clicker to control sequential presentation of each fiducial marker displayed on a standard PC monitor.

5.4 Software

The .NET 2.0 framework and WPF runtime were installed on the Dell XPS portable computer. The AR software developed using the WPF framework (see chapter 4) was utilised to augment fiducial markers with 3D media in real time.

Static Condition - Users were static throughout the test. Presentation of 3D media was achieved by displaying the visual fiducials on a computer monitor placed in front of the user. Presentation of media was controlled by the user simply clicking through a slideshow of twelve markers.

5.5 Stimulus Objects

The recognition task required that an array of twelve easily and instantly distinguishable objects be chosen. The first design choice therefore was whether to develop objects for recall in 2D or 3D. A 2D representation of an object is currently the most common



FIGURE 5.3: Sample of Highly Textured Geometry Created and Rendered in 3D Studio Max.

approach to presenting information in computer interfaces, an attribute of 2D is that object can be defined by a few simple geometric primitives. However, the predominately iconographic properties of 2D representations do not benefit from many salient attributes that help codify human recognition in the real world. By representing virtual information in 3D recall association is aided by offering enhanced colouring and shading, increased angular visibility and greater particulate density (see chapter chapter 3, section 3.4.1). It is reasonable therefore to suggest that the more redundant dimensions that are available to the user, the greater the chance of which to choose a feature on which to relate association. Since AR is primarily concerned with enriching the real world experience with virtual representations of real world objects, the use of 2D objects may propose a contradiction in terms. After deciding that 3D objects were more suitable than 2D, the choice of objects to use had to be decided. A common experiment (Kiries game) involves taking several everyday objects and laying them out on a surface. The user is given a set period of time to commit each object to memory and then has to recall the object and its position over a set time afterwards. The experiment implemented here shares clear similarities and was designed in kind, choosing a set of common objects that users could recognise instantly and therefore commit to memory more easily.

3D objects were developed in 3D Studio Max. Each object was highly textured whilst keeping the polygon count modest to ensure computational efficiency was not compromised. Each object was designed in order to promote geometric primitives to ensure objects could be identified quickly, for example sharp edges on a knife blade and an elongated ridged handle; finned wings and elongated spherical body on an aeroplane, pronounced tracks and turret on a tank (see figure 5.3).

5.6 Experiment 1A

5.6.1 Design

This experiment was designed primarily to investigate whether the addition of a frame of reference improved human performance. If so, this would suggest that a frame of reference is a determinant factor in regards to human performance and thus supports the benefit of employing a mixed reality technology such as AR to represent virtual information. The experiment needed to be designed to have an assessable outcome and would ideally require no previous experience from the participants. After talks with a colleague working with the Leicestershire Crime Scene Examination Department, an analogy was drawn between this study and crime scene examination. The findings here are especially compelling because the experimental conditions directly relate to currently employed methods in evidence review. In terms of the movement condition, the AR Off state can be associated with a situation where a forensic investigator may layout all the real evidence around a platform; the AR On state can be thought of as a combination of the previous analogy combined with a VR representation of crime scenes where evidence can be reviewed with spacial correlation provided by having a frame of reference to an objects location. Also the static conditions are very similar to how photographs of evidence might be reviewed sequentially on a computer screen. Obviously the selection of items (see table 5.1) would be unlikely to feature at a typical crime scene. However, we required items that would be recognisable by our participants (undergraduate engineers) and not require any specialised knowledge for interpretation and analysis.

TABLE 5.1: A Table to Show the Synthetic Objects Used in User Trials

Hand Gun	3.5" Floppy Disk	Tank	Halloween Pumpkin Head
Great White Shark	Sharp Knife	Human Action Figure	Fairy Tale Castle
Football	Dinosaur (T-Rex)	Commercial Aeroplane	Motorbike

Since this was a repeated measures experiment which required each condition to be conducted many times on a number of participants it was deemed inappropriate for testing on actual forensic investigators. Simple logistics concerning forensic professionals availability and even the sourcing of enough participants proved a repeated measures experiment of this magnitude inappropriate for the intended user demographic. It was

decided that the parameters of the intended experiment could be applied to a more available user group and still provide relevant data and this led to the following experiment design: This experiment compares human performance in four conditions;

Movement Condition (MC) - AR On Real world frame of reference is available and the user controls the viewable object by moving to each subsequent marker. Thus, spatial cueing and real world visual context was available.

Movement Condition (MC) - AR Off The user controls the presentation of virtual information in the same way as the AR On condition except they are simply presented in null space with no real world frame of reference available. Thus, spatial cueing, due to movement, was available but real world visual context was not.

Static Condition (SC) - AR On the user sits still, facing in one direction for the duration of each test manually controlling the presentation of each object sequentially and completes a series of repeated measures tests. Thus, no spatial cueing was available however, a real world visual context was available although this was effectively a static fixed image throughout each AR On test.

Static Condition (SC) - AR Off the user sits still, facing in one direction for the duration of each test manually controlling the presentation of each object sequentially and completes a series of repeated measures tests. Thus, no spatial cueing was available nor was any real world visual context.

Each subject is presented with twelve high resolution textured 3D objects that are individually assigned to fiducial markers. The practical implementation of the conditions was indifferent for AR On and AR Off. In AR On a webcam was used to present a window on the world in respect to the fiducial markers, AR Off would therefore be achieved by removing the webcam video stream so that each object would appear exclusively in the participants display (see figure 5.4). In the static condition the user still observes the world via the HMD but rather than present the markers in situ around the user they are presented on a computer monitor, the user 'moves' to the next object with a wireless clicker. To ensure consistency in the results each condition was performed over a series of repeated measures under all conditions with each participant. With each rendition

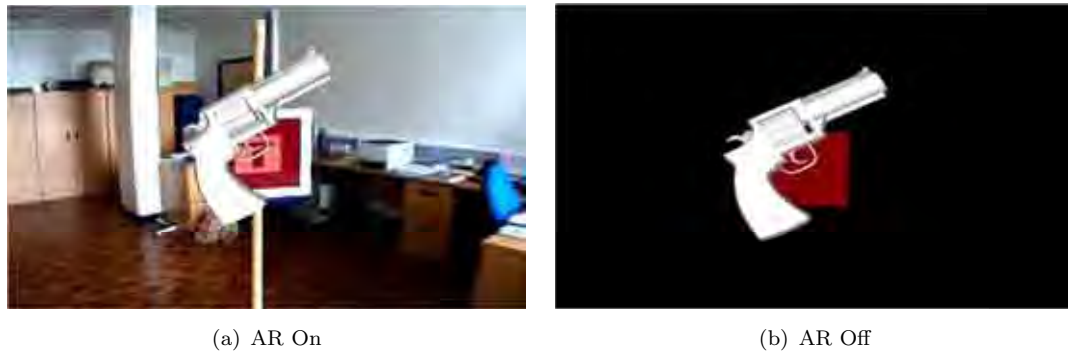


FIGURE 5.4: Comparison of relative conditions AR On and AR Off.

of the experiment the marker to which each 3D object was assigned changed to ensure any memory transfer between experiments was avoided.

The static condition was designed such that each marker was presented sequentially, ie the previous object could not be reviewed again. This could be construed as unfair since in the movement conditions users could simply move or look such to re-view an object. In light of this provisions were put in place to maintain as much transparency as possible between conditions. Users were instructed specifically to move about a clockwise-only direction and markers were spaced outside the viewable FOV, hence only one marker could be reviewed at a time.

5.6.2 Hypotheses

In order to test the research statement stated earlier, the following two hypotheses were formulated;

- a) AR will improve human performance via facilitation of an environment stabilised frame of reference (ESFoR). Such improvement is tested through comparison of AR On and AR Off conditions.
- b) AR will improve human performance via facilitation of a body stabilised frame of reference (BSFoR). Such improvement is tested though comparison of the movement condition and the static condition.



FIGURE 5.5: User Trial Scene Shot - User rotates clockwise about a horizontal plane viewing objects in sequence.

5.6.3 Procedure

Participants were given an introduction to augmented reality and briefed on the procedure of the experiment with specific instructions to report any physical or visual discomfort experienced. They were informed that they could withdraw at any time or could refuse permission for their results to be used, following the explanation, participants signed a consent form (see appendix E). Participants were seated on a standard office chair, with seat height adjusted so that the user could sit with their feet comfortably placed on the floor and their back straight. The seat was positioned approximately 1.5m from the markers (which were arranged equidistantly in a circle around the seat), as illustrated by figure 5.5. To ensure accurate registration of objects in all conditions a redundant test was carried out with each participant where the user would be asked to verbally announce their immediate interpretation of each object. For example it was preferable to allow a user to adopt and use their instinctive first interpretation of an object eg. shark - fish, football - ball, castle - house, pumpkin - monster. This was thought to promote a users natural cognitive inference in the recall stage.

Movement Condition The twelve markers were placed sequentially and equidistant around the user over 360 degrees at per-clockface intervals. The participant observed each marker in turn rotating through 360 degrees in a clockwise only direction. A total period of no more than sixty seconds was given to observe all twelve markers upon which a recall sheet was given (see Recall Sheet - Experiment 1A, appendix F) and another sixty seconds was allotted for users to record as best they could each object's explicit location relative to each marker. In accordance with standard practice of repeated measures experiments, the order in which conditions were executed (AR On, AR Off) was alternated. The user completed a total of 30 tests counter-balanced between each condition.

Static Condition Each of the twelve markers were represented on a slideshow, one marker per slide. The user navigated each slide in sequence via a wireless clicker, they could only review each slide once. A total period of no more than sixty seconds was given to observe all twelve markers upon which a recall sheet was given (see Recall Sheet - Experiment 1A, appendix F) and another sixty seconds was allotted for users to record as best they could each object in the correct order upon which they appeared. The user completed a total of 30 tests counterbalanced between each condition. The static condition trial was conducted with the same participant (within-subjects) group two weeks after the movement condition trial, doing such was believed to remove any memory effect from the movement condition to the static condition. To ensure satisfactory user compliance the practice session was re-run for the static condition also.

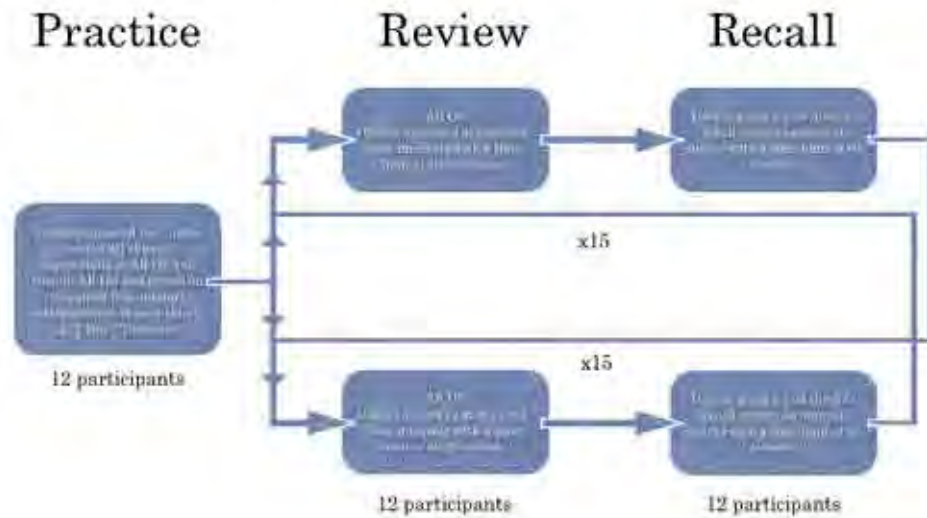


FIGURE 5.6: Design of study showing how the tasks carried out with each participant in the movement condition. The testing sequences of the AR On and AR Off were counterbalanced between participants

Each user wore a video based HMD device with a webcam affixed to the front, the HMD was connected to the Dell XPS portable computer running the fiducial marker recognition software. The subjects completed a practice run to confirm they were able to successfully interpret each object, they then completed a series of repeated measures tests in each condition. After each trial run, participants were given sixty seconds to recall items from memory to a test sheet.

Practice In order to ensure good user compliance throughout the trial, users were given a practice run in which they were given sixty seconds to review all twelve objects with the request to verbally announce their first interpretation of each object. This allowed the examiner to create an object interpretation key for each participant eg. participant 1 interpreted the 'great white shark' as a 'fish' and the 'fairy tale castle' as a 'house'. They were then given a further sixty seconds to recall all items from memory to a test sheet. Users were then asked if they were comfortable with

the test procedure and happy to continue, if necessary minor adjustments were made such as positioning of HMD and headgear before the test continued.

Review In the review section users reviewed all twelve objects pausing between objects in order to commit each one to memory. *Movement Condition* - In a counterbalanced trial order of AR On and AR Off over a total of 30 repeated measures tests, users reviewed all twelve objects by moving in a per-clockface fashion about 360 degrees in a clockwise only direction. During each test the user attempted to commit each object to memory. *Static Condition* - In a counterbalanced order of AR On and AR Off trials over a total of 30 repeated measures tests, users reviewed all twelve objects sequentially using a wireless clicker. During each test the user attempted to commit each object to memory.

Recall In the recall section users were presented a test sheet immediately after completing each test run. The test sheet was designed to complement the spatial layout of the virtual objects (see see Recall Sheet - Experiment 1A, appendix F).

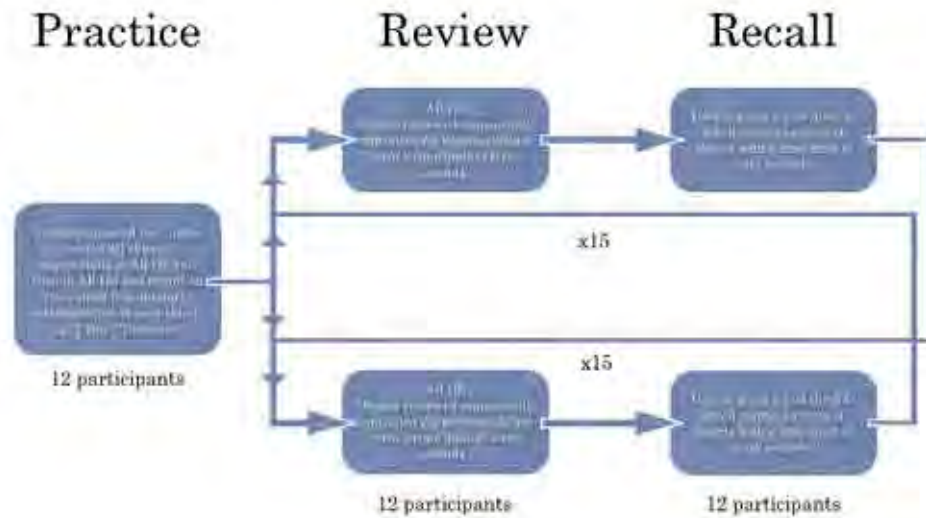


FIGURE 5.7: Design of study showing how the tasks carried out with each participant in the static condition. The testing sequences of the AR On and AR Off conditions were counterbalanced between participants

5.6.4 Results

5.6.4.1 Measurements

The first two tests were assigned to the practice session and used for familiarisation purposes only. Thus, those results were discarded from analysis. A total of 360 sets of data were collected from each experimental condition (12 subjects x 2 methods x 15 trials). The maximum score for a single data set was twelve, any ambiguity in a recording would be regarded as incorrect recall, such as 2 answers in the same box. The data were segregated for analysis into 3 discrete sets; primacy, mid-range and recency. A first blush analysis of the results displayed clear performance differences in recall between the 3-4 item and 9-10 item, this was interpreted as an indication of primacy and recency memory effects. Hence, primacy recall was concerned with items 1-3, mid-range 4-9 and finally recency 10-12. Results were subjected to a 3-way Repeated Measures Within-Subjects Analysis of Variance (using SPSS) for Position (12 Levels), AR (2 Levels)

and Motion (2 Levels). Degrees of Freedom (DF) are taken as Spherically Assumed for both the measure and its error, i.e $[F(df_{variable}, df_{errorvariable}) = Fvalue, p = pvalue]$ (where p is stated as 0.000 then $p < 0.001$). Interaction effects were explored by comparing respective means in primacy, mid-range and recency. Also to explore how task approach may have differed with the addition of movement further metrics are calculated to highlight the difference in recall between primacy and the remaining set (items 4-12) and also between recency and the remaining set (items 1-9). Comparison between recency recall and the remaining set (items 1-9) for Movement and Static Conditions alike highlight how task approach may have differed when users were able to move. The same such comparison between primacy recall and the remaining set is also calculated to highlight a similar effect.

Finally if a learning was found over the duration of the experiment condition it could be argued that this presented the possibility of a confound between the static conditions being run 2 weeks after the movement conditions. The presence of a learning effect was measured by comparing the mean recall over the first six and final six trials in each condition.

5.6.4.2 Analysis

TABLE 5.2: A table to show the average serial position recall in primacy, mid-range and recency stages and also the total average recall for items 4-12, items 1-9 and all 12 items in each experimental condition. Mean value calculated from a maximum score of 15 at each object position. Bracketed values refer to 1 SD distribution of data.

Condition	BSFoR	ESFoR	Primacy	Mid-Range	Recency	Items 4-12	Items 1-9	Total Mean Recall
MC AR On	YES	YES	9.69 (2.86)	2.85 (1.58)	7.86 (3.23)	4.52 (3.27)	5.13 (3.85)	5.81 (3.88)
MC AR Off	YES	NO	8.30 (3.59)	2.03 (1.77)	6.94 (3.52)	3.67 (3.40)	4.12 (3.89)	4.82 (3.65)
SC AR On	NO	YES	7.92 (2.77)	1.875 (1.64)	10 (3.67)	4.58 (4.58)	3.89 (3.53)	5.42 (4.43)
SC AR Off	NO	NO	7.94 (2.62)	2.40 (1.75)	9.11 (4.03)	4.64 (4.18)	4.25 (3.34)	5.47 (4.10)

Referring to the movement condition, table 5.2 shows how recall performance under the MC AR On condition compares with the MC AR Off condition. Under both conditions primacy (recall of the first three items) and recency (recall of the terminal three items) effects are particularly strong with primacy showing figures of 65% and 55% total mean recall for MC AR On and MC AR Off respectively. Recency is slightly less impressive

with figures of 52% AR On and 46% AR Off. Figure 5.8 illustrates graphically the discrepancy between each condition showing clearly that with movement AR On improves recall performance and demonstrates that recall is strongest in the primacy stage (position 1-3) and recency stage (positions 10-12). The data also suggests that differences between factors had an effect on task approach. In terms of recall of primacy items compared with the remaining set (items 4-12), primacy recall is a more dominant factor in the movement condition. MC AR On Primacy recall performance is 9.69 vs 4.52 mean recall for the remaining set, giving 114% improved recall in the primacy domain. Likewise, the same comparison in MC AR Off gives 126% better recall of primacy items. Comparing this with the static condition, likewise comparison gives 73% better recall of primacy items for SC AR On and 71% better recall of primacy items for SC AR Off. However, such comparison with recency items sees that recency recall is a more dominant factor for the static condition. SC AR On Recency recall performance is 10 vs 3.89 mean recall for the remaining set (items 1-9), giving 157% better recall of recency items. Likewise, the same comparison in SC AR Off gives 114% better recall of recency items. Comparing this with the movement condition, likewise comparison gives 53% better recall of recency items for MC AR On and 69% better recall for MC AR Off. What is clear from this analysis is that overall the serial position effect seems more dominant in the static conditions with a total percentile difference of 415% (157+114+73+71) for static condition metrics vs 362% (114+126+53+69) for movement condition metrics. Finally, a further comparison in regards to the effect of movement on performance is to look at its effect on overall performance. Total mean recall for MC AR On of 5.81 vs 5.42 for SC AR On, sees a 7% improvement with the addition of movement. The same comparison for MC AR Off vs SC AR Off, sees a 14% reduction in performance with the addition of movement. What is clear is that the addition of movement and/or AR does effect how the task was performed.

TABLE 5.3: Experiment 1A, 2x2x12 ANOVA Results Table.

Source	DF	Mean Square	F-Value	P-Value	Power
Position (1-12)	11(121)	680.871(2.408)	282.747	0.000	1.000
Motion (ON/OFF)	1(11)	3.361(19.895)	0.169	0.689	0.066
AR (ON/OFF)	1(11)	27.563(7.286)	3.783	0.078	0.427
Position x Motion	11(121)	24.111(2.236)	10.783	0.000	1.000
Position x AR	11(121)	3.782(1.956)	1.934	0.041	0.871
Motion x AR	1(11)	43.340(8.018)	5.405	0.040	0.564
Position x Motion x AR	11(121)	2.219(3.149)	0.705	0.732	0.372

The 3-way Repeated Measures Analysis of Variance (AR x motion x position of target, see table 5.3) shows a main effect of position of target [$F(11, 121) = 282.747, p < 0.0001$]. Although there is no main effect of Motion or AR, there is however significant interactions between Motion and AR [$F(1, 11) = 5.405, p = 0.04$], Position and Motion [$F(11, 121) = 10.783, p < 0.001$] and Position and AR [$F(1, 121) = 1.934, p = 0.041$]. The interaction in position is clearly explicable with reference to the distinctive and expected primacy and recency curves shown in figure 5.8 however, the interaction between Motion and AR requires further investigation.

To explore the interaction between AR and motion further, graphs comparing each condition state were plotted and analysed:

Movement ON - Compare AR On vs AR Off (see figure 5.8).

Movement OFF Compare AR On vs AR Off (see figure 5.9).

AR ON - Compare Movement On and Movement Off (see figure 5.10).

AR OFF - Compare Movement On vs Movement Off (see figure 5.11).

The interactions in the results imply that the presence of the video feed whilst ostensibly irrelevant to the task provides incidental contextual information that improves short term memory. This is supported through comparison of respective means in each experimental condition. Figure 5.8 clearly shows a reasonable improvement in average recall performance at each object position with an overall mean recall improvement of 0.99 items (see table 5.2). Comparing this trend to that shown in figure 5.9 the results suggest that when items are presented sequentially in the same spatial location the advantage for AR is removed. By contrasting the movement condition results to those in the static conditions it can be seen that spatial distribution of targets appear to improve primacy (9.69 and 8.30 against 7.92 and 7.94 mean recall for MC and SC respectively) although this advantage is lost in recency as the presentation of items in the same location leads to better recall of this portion (see table 5.2). Speculatively, the spatial presentation of items as observed in when a BSFoR is provided may counteract the decay of primacy.

The findings suggest that even where the real world is not task-relevant by offering richer cues to spatial position, participants were perhaps helped to orient their own

spatial frame of reference and/or extra incidental contextual information that improved the encoding of object identity (e.g., the synthetic tank is by the real plantpot).

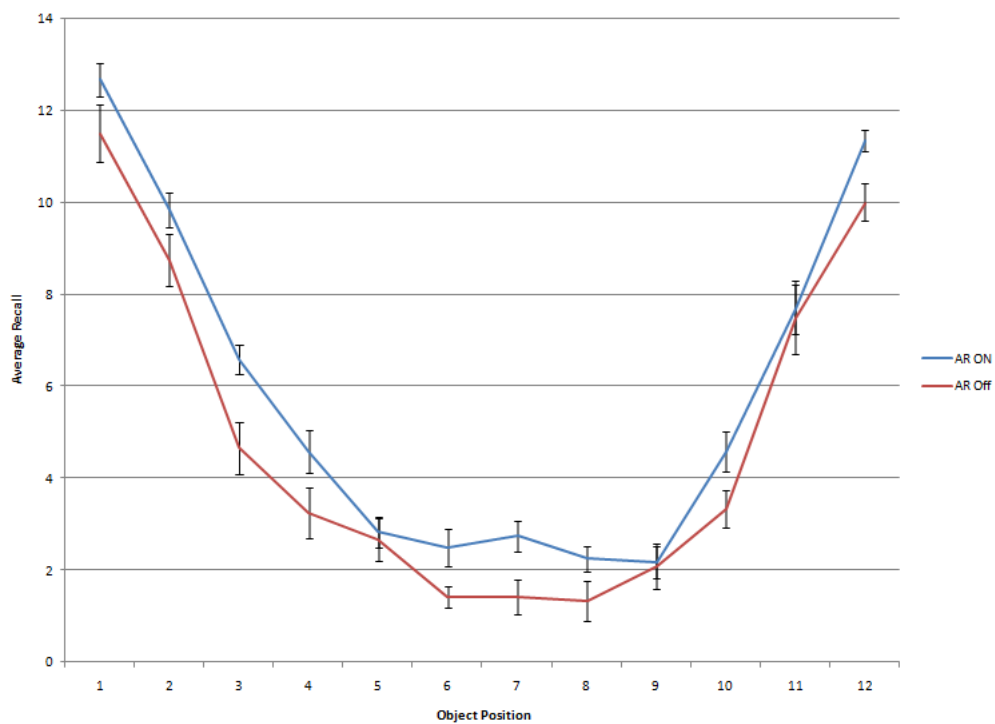


FIGURE 5.8: Graph of Recall Performance vs Serial Object Position for AR On and AR Off in the movement condition. The improved performance in the AR On condition show that the advent of a real world background whilst moving improves performance. Standard Error (SEM of 1SD).

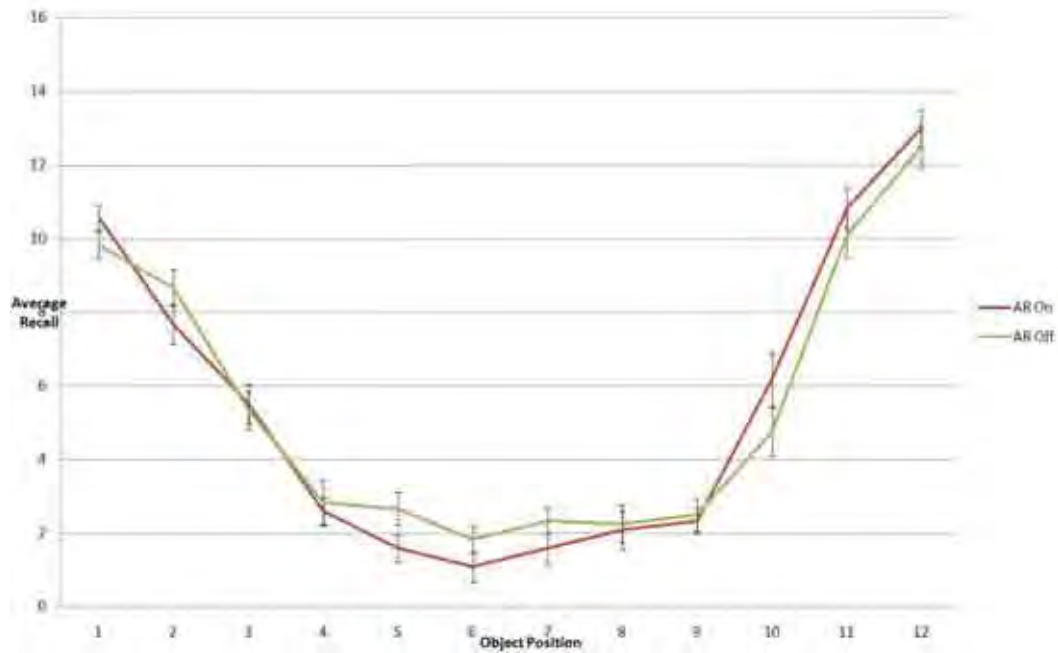


FIGURE 5.9: Graph of Recall Performance vs Serial Object Position for AR On and AR Off in the static condition. The lack of improvement in the AR On condition implies that the advent of a real world background alone did not improve recall performance. Standard Error (SEM of 1SD).

The results figures show clearly a serial position effect taking place in each experimental condition. Despite the obvious statistical tenuousness of collapsing positional data into one average recall figure it presents a good means to display the interaction effect of motion. Figure 5.8 shows clearly that the addition of AR when motion is employed improves recall. When motion is removed (see figure 5.9) however, the advantage for AR is lost. The recall figures (see table 5.2) show that total mean recall suffered most when motion was employed without the addition of AR. An interesting subtext to this experiment then, was to compare motion alone. Figures 5.10 and 5.11 were plotted to explore this effect. Referring to figure 5.10 it appears that without movement recall is improved in the recency domain but suffers in the primacy and mid-range. For AR Off however, it appears that the addition of motion impairs recall in the mid-range and recency domain and invokes no discernible difference in primacy recall. Thus, in the absence of AR performance suffers as a result of motion (see figure 5.11). It appears therefore, that motion interacts with AR positively but when AR is not being used it acts to distract the user from the task. This insinuates that movement itself does not aid recall. It is the relationship between movement and AR that improves recall.

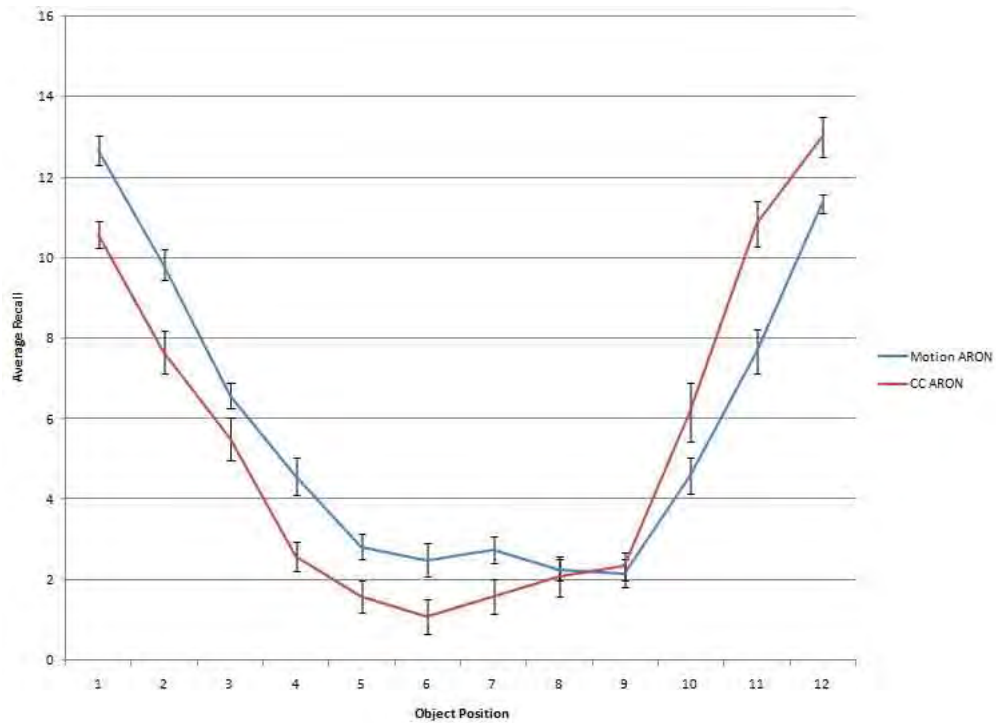


FIGURE 5.10: Graph of Recall Performance vs Serial Object Position for AR ON, Motion On and Motion Off. Referring to the recency items (10-12) it is clear the static condition improves recall. Standard Error (SEM of 1SD).

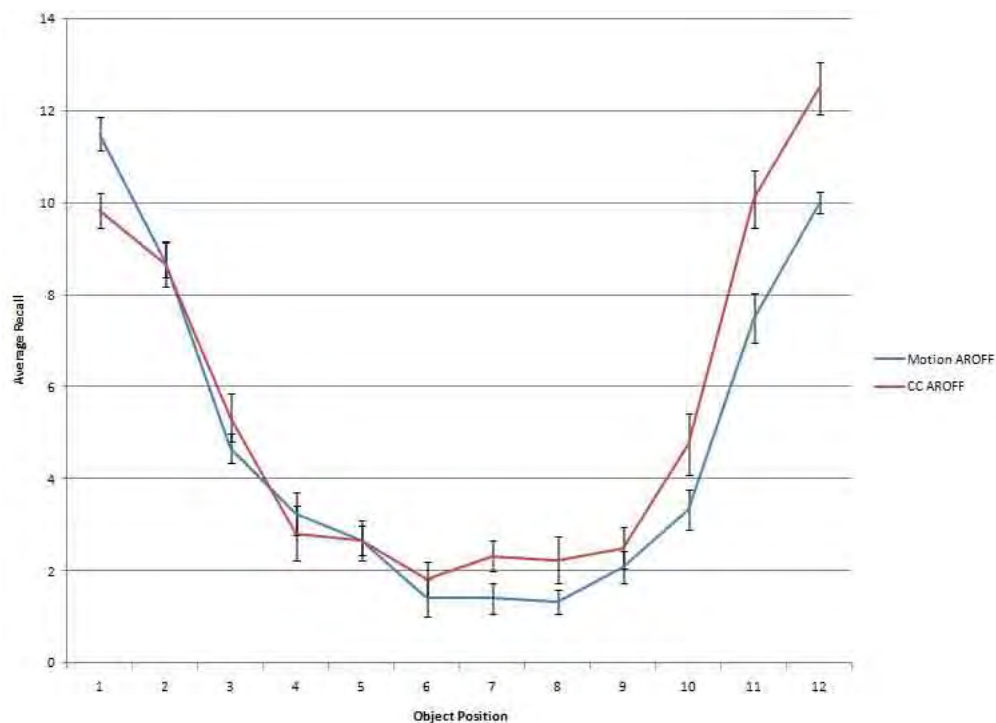


FIGURE 5.11: Graph of Recall Performance vs Serial Object Position for AR OFF, Motion ON and Motion Off. The graph shows clearly that for the most part, performance improves when motion is removed. Standard Error (SEM of 1SD).

TABLE 5.4: A table to show the mean total recall for trials 1-6 and trials 25-30 under each experimental condition. Mean value calculated from a maximum score of 12 for each trial run. Bracketed values refer to 1 SD distribution of data.

Condition	Trials 1-6	Trials 25-30
Static	4.347(0.826)	4.319(0.533)
Movement	4.403(0.396)	4.236(0.300)

Referring to table 5.4 for there to be a learning effect one would expect that over a subsequent number of trials a users recall performance would gradually improve. Difference in recall between the former and latter trials show no impact to recall performance in the static condition and negligible impact of -0.056 and -0.167 for the static and movement conditions respectively.

5.6.5 Discussion

In the static condition twelve participants sat in front of a monitor and observed each marker sequentially using a wireless clicker to move on to the next object. By testing the findings of the static condition under both AR On and AR Off scenarios, it was possible to ascertain that the real world reference frame was not providing undesired assistance to memory recall, such as semantic references or improved visibility of objects. The results from the static condition support this by showing that there is no discernible difference in recall performance when the background image is on or off if the user is stationary.

The results of the movement condition suggest that human performance is improved by providing a real world view as a frame of reference to the object in question. Such a reference frame is inherent in any AR based system and the results of this experiment support AR in the sense that it improves human performance. There are however, a few points to consider regarding this claim, namely that AR does improve performance in the context of the experiment that has been carried out here but does not consider how AR might effect performance when other possibly more complex metrics of movement are employed. So while AR does improve performance in terms of providing a spatial frame of reference this experiment does not explore how factors of movement present in a fully realised ergo, fully mobile AR platform effects human performance.

A further deduction to be made is that the interaction effect of position x motion implies that moving changes the users' ability to recall targets and as a consequence implies that the behavioural approach to the task may differ. Therefore, although initially it may appear that utilising AR technology does not yield a huge improvement in average total recall performance the results show that the use of movement may help to alleviate the serial position effect by making the impact of primacy and recency performance less severe. Since differences in average recall between movement and static conditions is non-significant the reduced SPE is more than likely the effect of behavioural approach and not any boundaries presented by the technology. This difference in approach is highlighted when comparing the recall improvement between primacy vs mean recall of the remaining set (items 4-12) and recency vs mean recall of the remaining set (items 1-9). As such, in the static condition the overall differences in recall performance are notably higher compared with the movement condition indicating that the addition of movement helps alleviate the serial position effect.

Although it is clear that the combination of a BSFoR and an ESFoR do improve human performance, it may be fair to summarise that there is overwhelming evidence to suggest that primacy and recency recall is the governing finding of this experiment. However, as implied earlier recall approaches between the movement condition and the static condition differ and this is supported by comparing conditions in terms of movement only; mean recall sees the addition of movement improves performance with AR On but sees reduced performance for AR Off (refer to table 5.2). This suggests that it is not motion per se that affects performance but only when considered in terms of the stimuli location (position) or the medium (AR On / Off).

Referring back to the initial hypotheses, it appears that the factors of a BSFoR and an ESFoR are co-dependent. The results showed that an orientational movement (body stabilised) only based frame of reference adds no value to recall performance until supplemented with a visual (environment stabilised) frame of reference. It would seem then that AR does improve recall performance because it inherently supports a body based 'moving' frame of reference to the surrounding environment.

It could be argued that a possible confound to performance exists due to the static conditions being run two weeks after the movement condition. The experiment draws on current memory theory, utilising short term memory recall as a suitable performance

measure. An important factor in the experimental design is that recall performance is subject to the parameters of the experiment and not skewed by a practice effect caused by doing a high number of repeated measures. If a learning effect were to be observed in transfer from a trial run that ran two weeks later this would imply that ability in short term memory recall improves with practice. However, if this were the case improvement would be observed within each run of the experiment ie one would expect performance in trials 25 - 30 to be better than trials 1 - 6. Analysis of the results has shown this not to be the case with mean recall between trials 1-6 compared with trials 25-30 for the movement and static conditions respectively showing negligible difference. This reaffirms that user performance was consistent throughout the test and performance was affected by the differentiating factors in experimental conditions alone.

A clear caveat to this analysis however, is that by collapsing the data into a single mean value, there is a question of statistical ambiguity in making the assumption that all these values belonged together. One way of compounding this analysis is to use chunking. Chunking or blocking the data allows us to produce a graph with a smaller number of values and hence make a more succinct deduction in the validity in the findings.

5.7 Experiment 1B - Primacy and Recency Memory Strength: Blocking

5.7.1 Design

The results in Experiment 1A give clear indication that memory is best at recalling the first observed objects (primacy) and the last observed objects (recency). This effect of primacy and recency memory has been observed by Altmann [139]. This research examines various cognitive theories and concludes that primacy dominance is a logical consequence of the underlying memory theory. In previous work it has been shown that memory can be improved by blocking objects together, a phenomenon linked to chunking in short term memory [140]. While experiment 1A shows that overall human performance improves with a real world frame of reference, it is clear that the most compelling finding from these results is that memory is stronger in primacy and recency. A second experiment was therefore required to even out primacy and recency effects over a larger physical area so that the effect of a moving real world frame of reference

could be observed more clearly. The addition of blocking would also help explain the interaction effect observed earlier that motion was only beneficial to human performance if supplemented by AR.



FIGURE 5.12: For each marker 3D geometry is displayed in a blocked group of three.

The experimental conditions were preserved from the initial experiment. The same participant group was used from the first experiment, due to their unavailability one substitution had to be made, a 25 year old male graduate was replaced by a 24 year old female graduate. The experiment procedure and methodology was also unchanged to ensure the only parameter of difference was the effect of blocking objects. It was decided that the 12 objects would be separated into groups of three per marker (see figure 5.12) leaving a total of four markers positioned in an equidistant circumference around the user (see figure 5.13). This required the user to move to just four locations over 360 degrees as opposed to twelve, however at each location there would be three times as much data to recall. The blocking trail was conducted two weeks after the previous static condition study, this was believed to be sufficient enough to expel any possible memory affect from the previous experiment.



FIGURE 5.13: User Surrounded by 4 markers at per compass intervals, each marker displays a randomised block of three items.

5.7.2 Hypotheses

The addition of blocking to the movement condition is implemented such to reinforce the hypotheses.

AR will improve human performance via facilitation of a body based and environment stabilised frame of reference. The addition of blocking will require a greater magnitude of movement between stimuli. The user will therefore, be encouraged to use an environment stabilised frame of reference to supplement their body stabilised frame of reference.

5.7.3 Procedure

Practice In order to ensure good user compliance throughout the repeated measures testing process, users were given a practice run in which they were given sixty seconds to review all four blocks of three objects with the request to verbally announce their first interpretation of each object. This allowed the examiner to create an object interpretation key for each participant eg. participant 1 interpreted the

'great white shark' as a 'fish' and the 'fairy tale castle' as a 'house'. They were then given a further sixty seconds to recall all items from memory to a test sheet. Users were then asked if they were comfortable with the test procedure and happy to continue, if necessary minor adjustments were made such as positioning of HMD and headgear before the test continued.

Review In the review section users reviewed all twelve objects pausing between objects in order to commit each one to memory. In a counterbalanced trial order of AR On and AR Off over a total of 30 repeated measures tests, users reviewed all twelve objects by moving in a per cardinal compass point fashion about 360 degrees in a clockwise only direction. During each test the user attempted to commit each object to memory.

Recall In the recall section users were presented a test sheet immediately after completing each test run. The test sheet was designed to complement the spatial layout of the virtual objects (see see Recall Sheet - Experiment 1B, appendix G).

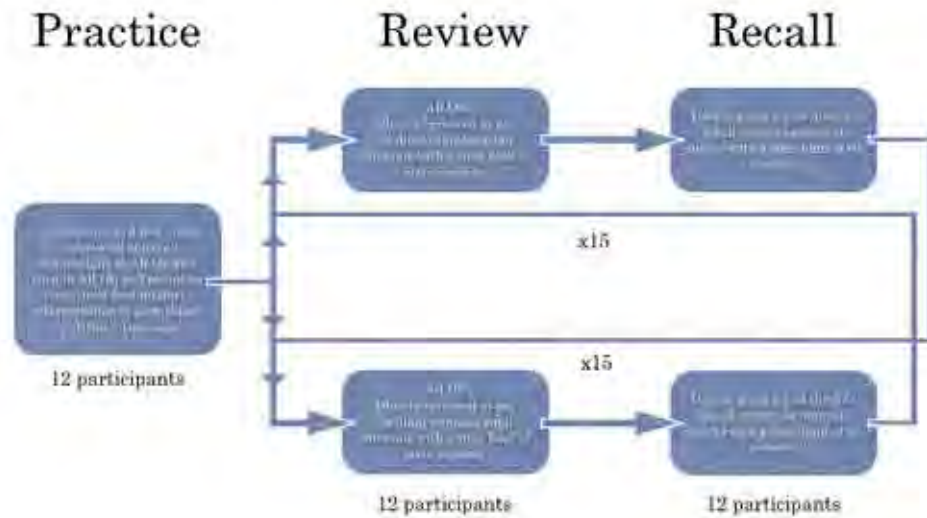


FIGURE 5.14: Design of study showing how the tasks carried out with each participant in the movement condition with the addition of blocking. The testing sequences of the AR On and AR Off conditions were counterbalanced between participants

5.7.4 Results

5.7.4.1 Measurements

The first two tests were assigned to the practice session and used for familiarisation purposes only. Thus, those results were discarded from analysis. A total of 180 sets of data were collected (12 subjects x 1 methods x 15 trials). The maximum score for a single data set was twelve, any ambiguity in a recording would be regarded as incorrect recall, such as 2 answers in the same box. The data was segregated for analysis into 3 discrete sets; primacy, mid-range and recency. The 12 objects were now blocked into four groups of three. Hence, primacy recall was concerned with items 1-3 (block 1), mid-range 4-9 (blocks 2 and 3) and finally recency 10-12 (block 4). Results were subjected to a 2-way Repeated Measures Within-Subjects Analysis of Variance (using SPSS) for Position (4 Levels) and AR (2 Levels). Degrees of Freedom (DF) are taken as Spherically Assumed for both the measure and its error, i.e. $[F(df_{variable}, df_{errorvariable}) = Fvalue, p =$

pvalue] (where p is stated as 0.000 then $p < 0.001$). Interaction effects were explored by comparing respective means in primacy, mid-range and recency.

5.7.4.2 Analysis

Figures 5.15 and 5.16 highlight an improvement in overall performance over the previous experiment, primacy, mid-range and recency recall is fairly improved and spread more linearly over the four object sets. Table 5.5 highlights a consistent improvement in recall in the AR On condition over the AR Off condition. Mean remembered items in the AR On and Off condition is improved with blocking. In the AR On condition, successful mean recall of 5.81 per position over 15 trials sees an improvement of 14.2% with an average mean recall of 7.94 items using blocking. Similarly in the AR Off Condition, successful recall of 4.82 items sees a similar improvement of 13.7% with a mean recall of 6.88.

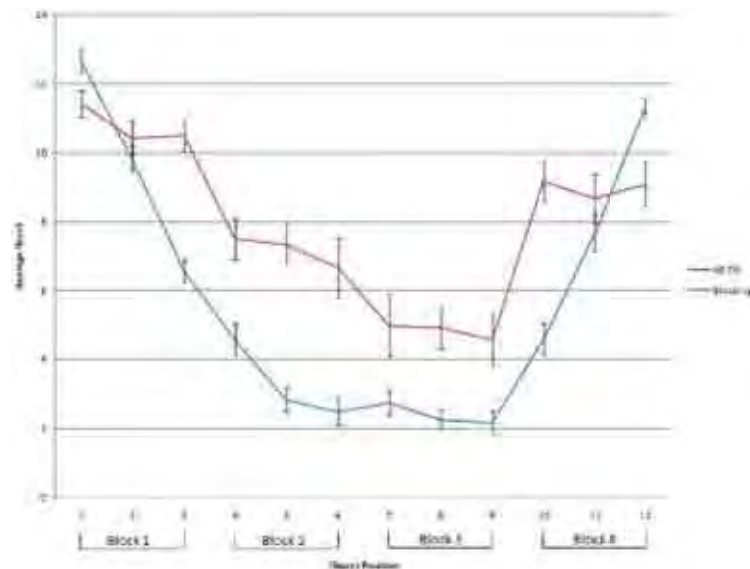


FIGURE 5.15: AR On condition experiment 1A compared with 1B (Blocking). Graph showing the additive affect in non primacy and recency by way of implementation of blocking objects together. Standard Error (SEM of 1SD).

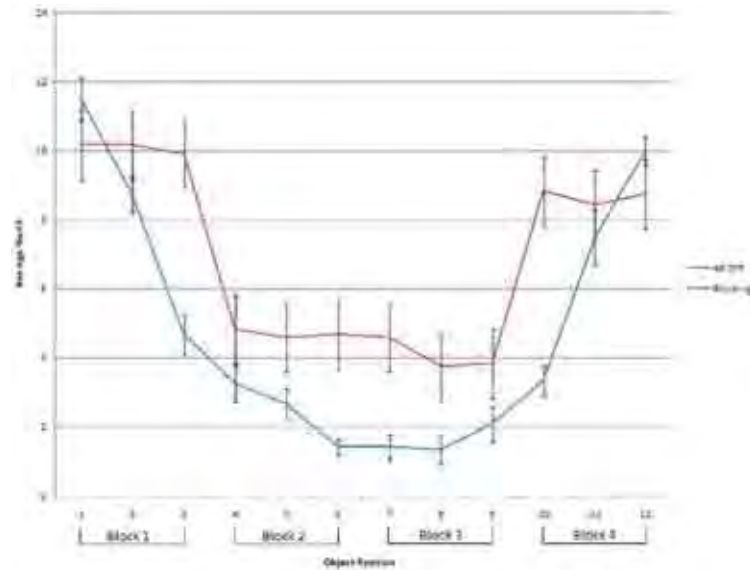


FIGURE 5.16: AR Off condition experiment 1A compared with 1B (Blocking). Graph showing in the additive affect in non primacy and recency by way of implementation of blocking objects together. Standard Error (SEM of 1SD).

TABLE 5.5: A Table to Show the Average Recall in Primacy, Mid-range and Recency Stages and Total Average Recall from a Maximum Score of 15 at each Object Position. For clear comparison of data the previous AR movement condition is compared to the results for the blocking condition. Bracketed values refer to 1 SD distribution of data.

Condition	Moving	Blocking	Primacy	Mid-Range	Recency	Total Mean Recall
MC AR On	YES	NO	9.69 (2.86)	2.85 (1.58)	7.86 (3.23)	5.81 (3.88)
MC AR Off	YES	NO	8.30 (3.59)	2.03 (1.77)	6.94 (3.52)	4.82 (3.65)
Blok AR On	YES	YES	10.78 (1.83)	6.00 (2.60)	8.97 (2.06)	7.94 (3.30)
Blok AR Off	YES	YES	10.08 (1.58)	4.38 (2.58)	8.67 (2.34)	6.88 (3.43)

Table 5.5 shows the improvement in recall performance with the addition of blocking. Interestingly recall of primacy and recency items sees less improvement using blocking for AR On than for AR Off. We observe 10% and 17.6% increases in recall in the primacy stage and 12% and 20% increases in the recency stage for AR On and AR Off respectively. However, it is in the mid-range where we observe a dramatic improvement in recall especially for AR On. Using blocking sees a 53% improvement for AR On and a 44% improvement for AR Off. The overall affect of using blocking sees an improvement to recall performance of 27% and 30% in AR On and AR Off conditions respectively.

Collapsing data sets by position results in 4 data points located at per cardinal compass points. Repeated Measures Analysis of variance (see table 5.6) shows a significant main

effect of AR (on/off) [$F(1, 11) = 103.571, p < 0.001$] and Position (1-4) [$F(3, 33) = 11.544, p < 0.001$].

TABLE 5.6: 2-Way ANOVA, Experiment 1 Blocking. Rounded to 3 s.f, bracketed values refer to error.

Source	DF	Mean Square	F-Value	P-Value	Power
Position	3(33)	92.066(7.975)	11.544	0.000	0.999
AR (On/Off)	1(11)	5415.010(52.283)	103.571	0.000	1.000
Interaction	3(33)	1.760(11.003)	0.160	0.922	0.076

5.7.5 Discussion

The results show that with grouping a significant improvement in human performance can be observed. This improvement looks to be compatible with the findings from experiment 1A. The most notable improvement is found outside the primacy and recency domain, as observed in figures 5.15, 5.16 for AR On and AR Off respectively. By presenting the twelve objects in groups of three at each cardinal compass point as opposed to having objects as per a clockface in the first experiment the governing effect of primacy and recency on the results is alleviated. The positive effect to human recall performance can therefore be attributed to the presentation of a real world background. Results showed that the least effective condition was AR Off where users were required to physically move about a 360 degree axis and observe markers without a background image. Analysis of the ANOVA results show there was no interaction effect between AR (On/Off) and Position when items were grouped in this way. In the blocking condition users moved to four positions rather than 12 and their degree of movement between stimuli therefore increased as such interpretation of each block of stimuli could be independent of the last. Further analysis of the data revealed that very significant gains were observed in mid-range recall (blocks 2 and 3) but not so much on primacy (1) and recency (4) recall. What this suggests is that where recall was much easier ie. primacy and recency users did not require the aid of an ESFoR, possibly because serial recall effects (first and last seen objects) was still prominent. However, where recall became much more difficult (ie. mid-range items) the need for additional stimuli from the ESFoR became a necessary aid to the memory encoding process.

This implies that items were referred to in reference to their in-situ position in the real world rather than each other (as was the case with 12 sequential positions) thus giving the user more spatial information of the real world to rely on when recalling the objects. Analysis in experiment 1A required further analysis of respective means to explore the meaning of the interaction between motion and AR. The blocking experiment supports these findings that the addition of a real world background (ESFoR, called AR On) whilst moving (BSFoR) improves performance. This evidence lends itself well in reference to the fact that the BSBFoR and ESBFoR that AR technology provides does have a determining effect on performance that improves recall.

Factors limiting currently available AR technology conjure pre-conceived notions that performance using mixed reality will have a negative effect on human performance. It could be argued then that a dependence on a visual background as an exo-centric reference to an objects position in real space is required more so because of the heavily diminished resolution of the objects being viewed through the HMD. A great deal of previous research however, does not support this argument. One such study Lemos et al. [141] showed that photo-realistic images were associated with a higher workload for image identification tasks, and additional visual detail did not increase performance for either identification or performance tasks. Research in earlier chapters showed that humans can perform well even where resolution is poor (see also, Ruddle and Lessels [75]). This begs the question *what changes to our view of the world affect human performance?* It is not merely the visual representation of the world that affects human performance but more importantly the extent to which the viewer is able to use their physical movement within a perceptible real world. This suggests that a key issue relates to the manner in which information sources (real and virtual) can be operated on in a similar manner. Moving along a single axis (horizontal, clockwise) allows convergence of the real and the virtual in this study. The role of the background information could, therefore, support this convergence.

Chapter 6

AR Performance Study - Immersive Reality vs Real Life

6.1 Introduction

The previous chapter was concerned with proving that a real world reference frame (inherent in any AR system) improved user performance. While this proves that the medium and context with which information is presented does affect human performance it does not make any comparisons to alternative ways of presenting information. In this chapter human performance over three information mediums across the MR continuum are compared. Real Life, Augmented Reality and Virtual Reality. The experiment utilises a prepared real life environment for which real objects are supplemented with virtually rendered renditions in the real environment to satisfy an AR condition and also a synthetic environment to satisfy a VR condition.

The AR system developed for the previous experiment was used. For the VR condition however, a new system was developed. This was achieved using a novel tracking interface, integrated into a popular games engine customised for use in this experiment.

This chapter aims to satisfy the following research statement:

Movement has been shown to be a governing factor affecting performance in AR. How does AR compare to other modalities such as; entirely synthetic environments and stimuli (virtual reality) and real environments and stimuli (real life).

6.2 Participants

Eight subjects each took part in five trials of each condition. All participants were university graduates aged between 21 and 30 years of age with an average age of 26. Of the eight participants 6 were male, 2 were female. None of the participants had taken part in any of the previous experiment trials.

6.3 Hardware Design and Software Development

In the previous experiment a binocular HMD was utilised however after some further user testing with the device, participants commented that the unit limited their head movement. The experiment in this section is concerned with facilitating and encouraging head rotation. Talks with a research fellow who received a doctorate researching the effects of counter-balancing the centre of gravity above the neck concluded that a heavy HMD coupled with tracking equipment could introduce undesirable offsets to the experiment proceedings, such as user fatigue and reduced head motion (see Knight and Baber [142]). For these reasons a monocular HMD approach was considered to be preferable as most affordable variations comprised a suitable form factor that could be modified easily and were low in weight.

HMD *Liteye LE-700A*

The Liteye LE-700A (see figure 6.1) is a lightweight (80 grams), low power device (0.4W. USB Powered). Resolution of the device is 800x600 with a FOV rating of 28°. The monocular design incorporates a fully adjustable head mount that can be adjusted to satisfy small human visual impediments such as a myopic condition. This is due to having diopter adjustments between +2 and -5. This means that in many cases the device can be used without the need for glasses.



FIGURE 6.1: LiteEye LE-700A SVGA Head Mounted Display Unit

Virtual Environment *Crytek CryEngine Sandbox SDK 1.3*

The CryEngine is the 3D graphics engine behind popular computer video game Far Cry released for Windows in March 2004. The engine is a 3D sandbox tool that allows developers to develop game levels using custom geometry or entities with custom effects and behavioral properties. CryEngine is robust in design and is supported by good documentation and an active development community. Many cutting edge graphics features are implemented and easily accessible to developers such as high quality vertex shading that utilizes DirectX 9.0c Shader Model 3.0, Polybump Normal Mapping and High Dynamic Range (HDR). Developers also benefit from a realtime sandbox environment. Other games engines such as Valves Source Engine (Half-Life), The Unreal Engine 3 or IDs Quake 4 SDK require the user to recompile and reload every time game content is manipulated or changed. To manipulate texture and model compilation with the Source Engine for example requires a high level of text-editor scripting with lengthy console commands. In the CryEngine geometry can be moved and resized within the environment, textures can be altered and modifications to effects such as lighting and shaders can be viewed in real time thus simplifying the design process considerably.

Using 3D Studio Max a workspace was documented then accurately modelled, photographs of the environment were used to create direct draw textures in Adobe Photoshop. Figure 6.2 shows the real scene and the equivalent photorealistic synthetic environment. The 3D Studio Max (3DS) geometries were imported into the CryEngine, textured and then coupled with dynamic light sources to create a compelling VR representation of the real test environment shown (see figures 6.3, 6.4).



FIGURE 6.2: Synthetic Renditions of Real World Objects Imported into the CryEngine



FIGURE 6.3: Real World Workspace for Experiment Trial



FIGURE 6.4: Equivalent 3D Virtual Representation of Real World Workspace

Headtracking *Nintendo Wiimote*

The Nintendo Wiimote is the novel control interface behind Nintendo's Wii games console. This use of technology offering a motion based approach in gaming has seen unprecedented commercial success particularly amongst groups not usually associated with computer gaming such as the female and older demographic. The Wiimote utilises a 3-axis accelerometer to register orientation in the pitch and roll domains. Since the accelerometers are not geo-stabilised using gyroscopes (see inertial trackers, chapter 2, section 2.4.5) an unconventional tracking approach is used. Housed in the front of the Wiimote is an optical sensor covered with an IR filter lens. The sensor is capable of reporting the limited information of up to four IR sources such as intensity and an enumerated number assignment and their relative locations. The location of each dot is calculated relative to the cameras FOV. The sensor returns values in the range of 0 - 1023 and 0 - 767 for horizontal and vertical planes respectively. Integer increments equate to 0.04° thus, FOV is perceived to be approximately 41° in the horizontal field and 31° in the vertical field. Calculations of IR sources relative to the camera are done on the Wiimote IC (Integrated Circuit). This is most probably because transmission of the cameras raw pixel data would introduce a prohibitive load on bandwidth usage. Communication with the Wiimote is via a Bluetooth wireless link designed to follow the bluetooth human interface device (HID) standard used by common wireless peripherals such as mice, keyboards and mobile phones. Authentication or encryption features of

the HID bluetooth standard are not required to communicate with the Wiimote. This makes it possible to write code to interface with the Wiimote on a PC with a compatible bluetooth adapter.

Once the Wiimote is paired with with a PC, it is identified as a HID-compliant device. Therefore, to initiate a communication dialogue with the device it is necessary to use the HID and Device Management Win32 APIs which are defined in the Windows Driver Kit (WDK). There is however, no built-in support for these APIs in the current .NET runtime. Fortunately API functions can still be called from within the .NET framework using P/Invoke to directly call methods within a Win32 API. This method is often called wrapping and requires that the right structure definitions and method signatures are used to correctly marshal data through to Win32 libraries. A series of steps can be undertaken to successfully communicate with any HID device. Each HID device is defined by a vendor and product ID, using the known product and vendor ID of the Wiimote we simply get a handle to the list of all devices contained within the HID class then enumerate to the correct product and vendor ID. Once found a filestream method is set-up to send and receive reports via an asynchronous callback buffer.

As with all popular 3D games engines orientation within the virtual environment is controlled using a mouse. This mouse-look interface can simply be controlled or emulated through code. Using the discreet positional information received from the Wiimote it was possible to emulate the behaviour of a mouse so that the Wiimote could control the environment in a way that emulated natural head rotation. This was achieved using a triangulation algorithm of the positions of IR sources placed at separated fixed distances from the FOV of the on-board optical sensor.

6.4 Stimulus Objects

In the previous experiment recall performance when using twelve distinct 3D objects proved to be quite low, alignment between mediums with such a large array set would be difficult and could have adverse effects on performance (see chapter 3, memory and search). In this study we sought to reduce this complexity and varied one dimension only, colour and limit this dimension to three levels. Furthermore we required objects that could be presented in real life or as synthetic virtual items in such a way as to be as

closely aligned as possible to their real counterparts, (see chapter 3, section 3.4.2). Hence the decision to use bottles, a simple granular object that was universally distinguishable seemed reasonable.

The experiment required that the user accept the objects in their virtual or real guise as being identical apart from one definable property, colour, thus simple small drinks bottles were selected. With the aid of a 3D scanning device, it is possible to create highly accurate 3D virtual representations of real objects. This process works best on simple geometry thus, the use of bottles allowed for a convincing virtual representation that when visualised through the HMD was not wholly indifferent to its real counterpart. Also on a practical application of the concept bottles are commonly associated with crime scene investigative work, often proving valuable for DNA scraping or finger print lifting.

6.5 Design

In order to satisfy the design requirements, the experimental procedure under each reality condition had to be indifferent. Such was achieved through the nature of the experiment design. Each experiment essentially required the same user interactivity and input.

The user was presented with twelve bottles each in one of three primary colours, the colour choice for each bottle was randomised between repetitions. The objects were presented over a flat surface that could be examined freely with three degrees of freedom (pitch, roll and yaw) over a set time period. The transfer between conditions is controlled such that each is approached in the same way. After the user has freely examined all twelve objects they were required to complete a recall task in which they try to the best of their ability to record the bottle colour in its correct position (see Recall Sheet - Experiment 2, appendix H).

The three conditions were identical by comparison in terms of object layout, hardware layout and experimental execution. The difference between the conditions was therefore restricted to factors inherent in the presentation medium being utilised.

Real Life Coloured Bottles that were used to model synthetic equivalents in 3D, are positioned over the fiducial markers such to simulate the other conditions, as shown in figure 6.5.

Augmented Reality Using the fiducial marker system each object is appended to a specific marker. The user sees the environment through a HMD WOW setup.

Virtual Reality Consistency between conditions is maintained by keeping all variables unchanged; the user occupies the same work space, uses the same equipment and performs the same task. The system therefore behaves as if the user was observing the same environment when in fact it is an entirely virtual representation of the real surroundings (see figure 6.4).



FIGURE 6.5: Real World Experiment Scene - figure demonstrates real bottles placed over markers, in the AR condition synthetic equivalents are presented in place of the marker.

Strict measures were undertaken to ensure each condition was equivalent. Between real and virtual scenarios, relative locations of bottles, bottle size and freedom of movement for the user had to be kept consistent throughout the trial. While the aim of the task for participants was to memorise the items and recall them. Recall performance was measured to ensure task compliance, memory was not the main focus of the investigation, rather it was used as a proxy task through which participant-chosen scanning of a simple scene could be examined. Owing to the physical nature of the AR/VR headset it was not

6				8
	2		4	
		1		
7				9
	3		5	
10		11		12

FIGURE 6.6: Grid shows distribution of stimuli over the scan area

possible to directly measure eye movements, so instead the targets were widely spaced out within the scene (see figure 6.5) and head movements were measured using tracking technology derived from the Nintendo Wii controller which tracked, pitch, roll and yaw from which angle could be calculated. The distribution of stimuli can be seen in figure 6.6.

6.6 Hypotheses

In order to test the research statement stated earlier the following hypotheses were formulated.

- a) There should be no discernible difference in recall performance between conditions. The trial is limited to a constrained environment that should encourage users to adopt the same strategy under each condition. Therefore information encoding should be uniform.
- b) Search strategy in each condition should be indifferent. In all three conditions the user's experience of the environment is constrained by the same limitations and technology.

6.7 Procedure

Participants were given a basic introduction to AR and its comparison to the more familiar conditions, VR and RL. It was requested that participants report any physical or visual discomfort experienced. They were informed that they could withdraw at any time or could refuse permission for their results to be used, following the explanation, participants signed a consent form (see appendix E).

Participants were seated on a standard office chair, with the seat height adjusted so that the user could sit with their feet comfortably placed on the floor and their back straight. The seat was positioned approximately 1.5m from the markers which were arranged in front of the user, as illustrated by figure 6.5. The distance from the scene was calibrated such that the user would be required to tilt their head across approximately 70 degrees horizontally and approximately 30 degrees vertically. All participants were briefed on the procedure of the experiment. Each user wore a monocular video based HMD device, the second eye was covered with a fully opaque eye patch. Pilot tests during the setting-up of the equipment showed that at most one could see two bottles from any fixation, but that when one bottle was held in central view the likelihood of seeing other bottles was minimised. Thus, we assume that a fixation on a bottle was most likely to result in the participant seeing only one bottle in their immediate FOV. The position of the bottles (and corresponding virtual environment) were modified until this criterion was met as well as practicable. Twelve markers were spread over a flat 8 ft x 10 ft (2.4m x 3m) vertical surface. For each test run, the participant was given sixty seconds to observe the environment with the instructions to look at all of the bottles on the wall and you will be asked to remember the color of the bottles at all twelve location. Following this, the participants turned away from the wall, lifted the head-mounted display and complete a recall sheet. There was a time limit of 60 seconds allotted to record the color of the bottles at each given location in order to force recall performance. During this time, the bottles were switched around by the experimenter, following a pre-defined schedule, so that each run would not involve the same bottles in the same positions.

The Wiimote tracking hardware developed primarily for the virtual environment (VR) condition was utilized in all conditions to track head movement and provide data for further analysis. This provided data on head-movement and an allowed for an index of dwell time to be calculated.

Participants were allocated to condition (AR, VR, RL) on order of appearance, i.e., participant 1 had AR first, participant 2 had VR first etc. The second condition was then selected from the remaining two on an alternating basis. Care was taken to ensure that exposure to the three conditions varied across all participants in a Latin Square design. Each user was only subjected to each predetermined configuration of objects once. Between conditions the configuration varied, ie no RL condition was similar to a AR or VR condition. The participant was given a period of no more than sixty seconds

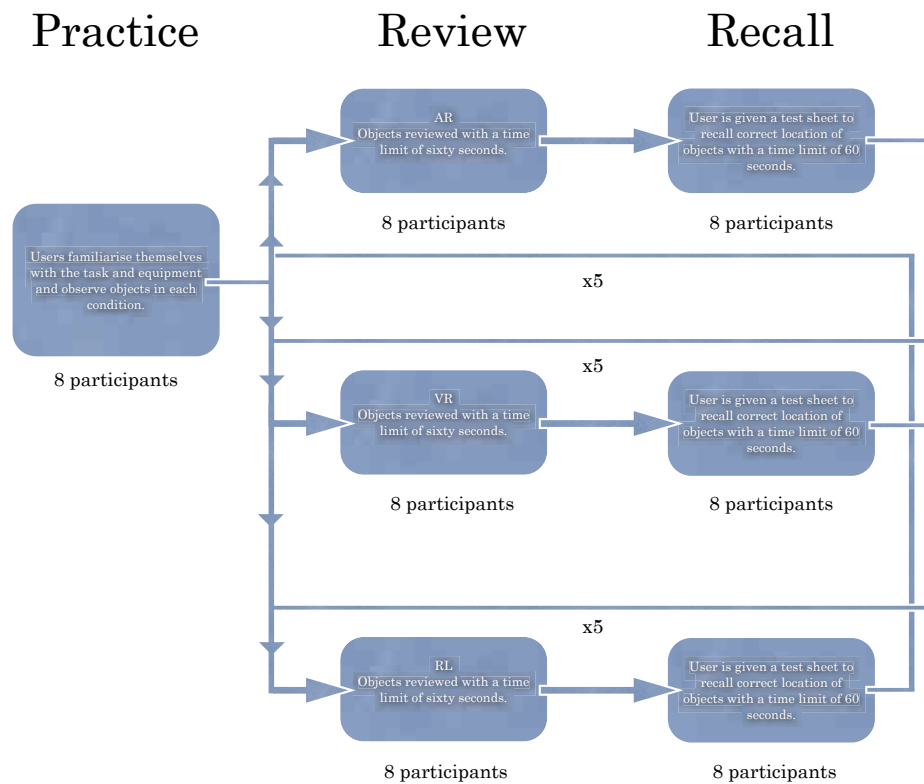


FIGURE 6.7: Design of study showing how the tasks carried out with each participant. The testing sequences of the VR, AR and RL conditions were counterbalanced between participants.

to observe and record to memory the colour of the bottle at each of the twelve locations. A further sixty seconds was then allotted to record the colour of the bottles at each given location. In each condition the experiment was repeated no less than five times. The Wiimote tracking hardware developed primarily for the VR condition was utilised in all conditions to track head movement and provide data for further analysis.

6.8 Results

6.8.1 Measurements

The first three tests were assigned to the practice session and used for familiarisation purposes only. Thus, those results were discarded from analysis. A total of 120 sets of data were collected from the experiment (8 subjects x 3 methods x 5 trials) which comprised of a within-subjects design. The maximum score for a single data set was twelve, any

ambiguity in a recording would be regarded as incorrect recall, such as 2 answers in the same box. Results from mean total dwell time and recall were combined at each object position for each condition (VR, AR, RL) and subjected to a Repeated Measures Within-Subjects Two-way Analysis of Variance (using SPSS) for Position (12 Levels) and Condition (3 Levels). Degrees of Freedom (DF) are taken as Spherically Assumed for both the measure and its error, i.e. $[F(df_{variable}, df_{errorvariable}) = Fvalue, p = pvalue]$ (where p is stated as 0.000 use $p < 0.001$). Post-hoc tests of significant main effects were performed using Fisher's PLSD (using SPSS).

6.8.2 Analysis

TABLE 6.1: Analysis of Variance Table for Dwell Time per position, 3x12 design. (Bracketed Values refer to Error)

Source	DF	Mean Square	F-Value	P-Value	Power
Position	11(77)	27.646(1.563)	17.688	0.000	1.000
Condition	2(14)	18.367(2.964)	6.197	0.012	0.812
Interaction	22(154)	2.270(1.243)	1.827	0.019	0.975

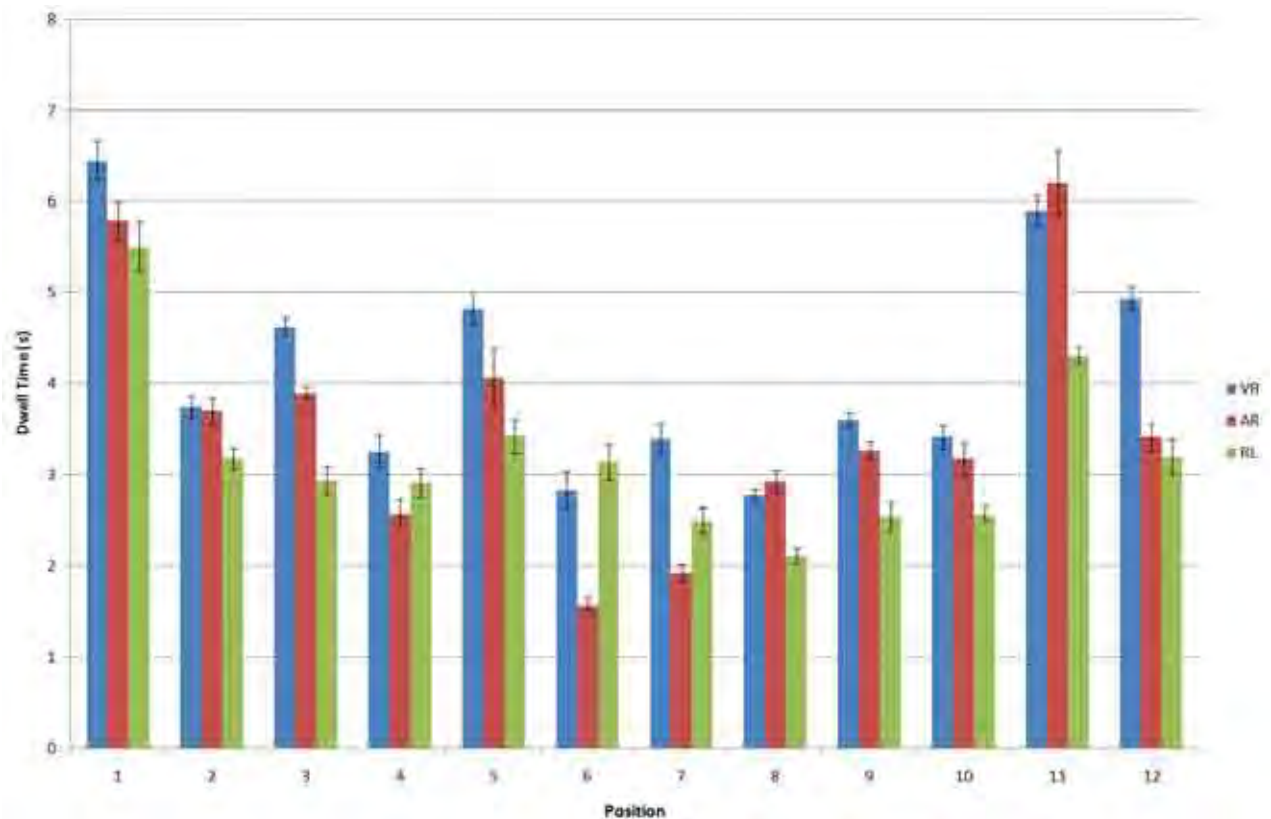


FIGURE 6.8: Dwell Time x Position for AR, RL and VR conditions. Standard Error (SEM of 1SD). Total Dwell Time, 49.15, 44.02 & 41.28 for VR, AR and RL respectively.

A two-way Analysis of Variance for position of target (12 levels) and condition (3 levels) revealed a main effect of condition on dwell time [$F(2, 14) = 6.197, p = 0.012$] and a significant interaction [$F(22, 154) = 1.827, p = 0.019$]. To explain the interaction post-hoc analysis was required. Post-hoc pair-wise comparisons, using Fisher's PLSD, indicate that there was no difference between the AR and RL conditions $p = 0.350$, but significant differences between VR and AR ($p = 0.015$) and between VR and RL ($p = 0.025$). There was also a main effect of position of target [$F(11, 77) = 17.688, p < 0.001$], which can be clearly seen in figure 6.8 with participants spending more time looking at central locations (1 and 11) than the other locations.

One can see some variation in the performance in the VR condition (compared with the other conditions), particularly in the lower parts of the grid (7, 9, 10, 11 and 12). Performance in the AR and RL conditions tend to mirror each other; although there is a slightly higher dwell time for AR on most locations, possibly as a result of need to ensure registration of the target, this did not appear to cause a measurable difference in performance.

TABLE 6.2: Analysis of Variance Table for Recall Performance per position, 3x12 design. (Bracketed Values refer to Error)

Source	DF	Mean Square	F-Value	P-Value	Power
Position	11(77)	8.034(0.463)	17.349	0.000	1.000
Condition	2(14)	2.358(2.695)	0.875	0.439	0.171
Interaction	22(154)	0.796(0.543)	1.447	0.101	0.920

These variations in dwell time also impacted upon recall performance, referring to table 6.3, recall in all three conditions appears to follow similar trends in terms of probability of recall at each position of target. This observation is reflected in a 2-way ANOVA for position of target (12 levels) and condition (3 levels) that demonstrates an expected main effect of position [$F(11, 77) = 17.349, p < 0.001$] but no main effect of condition and no interaction. A set of correlations between the probability of recall for any given item in each experiment condition correlated with dwell time, suggesting that the longer an item was examined, the more likely it was to be later recalled (see figures 6.9, 6.10, 6.11).

TABLE 6.3: A Table to Show the Average Recall from a Maximum score of 5 at each position. Bracketed values refer to 1 SD distribution of data.

Position	VR	AR	RL
1	4.75 (0.46)	5.00 (0)	5 (0)
2	3.625 (0.74)	3.88 (0.83)	4 (0.93)
3	3.5 (0.93)	3.63 (0.52)	4 (0.93)
4	3.5 (0.76)	3.88 (0.99)	3.88 (1.13)
5	3.5 (0.76)	3.13 (1.13)	3.88 (0.64)
6	3.125 (0.64)	3.00 (1.20)	3.38 (0.52)
7	2.875 (0.64)	3.63 (0.92)	3.00 (1.07)
8	3.5 (1.07)	3.5 (0.93)	3.00 (1.20)
9	3.125 (0.64)	3.13 (1.36)	2.88 (0.99)
10	3.25 (1.04)	4.38 (0.74)	3.76 (1.04)
11	3.625 (1.06)	4.63 (0.52)	4.86 (0.34)
12	4.5 (0.76)	4.50 (0.54)	4.38 (0.52)
Total	3.57 (0.93)	3.85 (1.04)	3.83 (1.04)

In each experiment condition the scan paths between items were recorded. This was achieved by analysing and collating the raw movement data from the Wii remote tracking device. Zones of interest were then declared and data mining techniques were used to map a scan path. Despite differing dwell times, the patterns in which targets were visited and the favoured repeated rehearsal paths taken around them were very similar

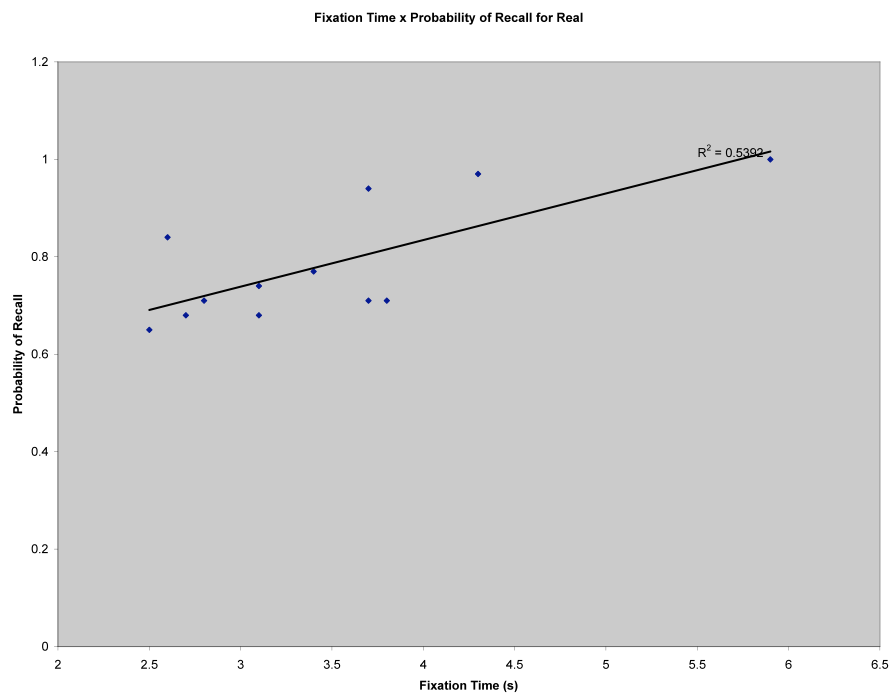


FIGURE 6.9: Real Targets: Fixation Time versus Probability of Recall.

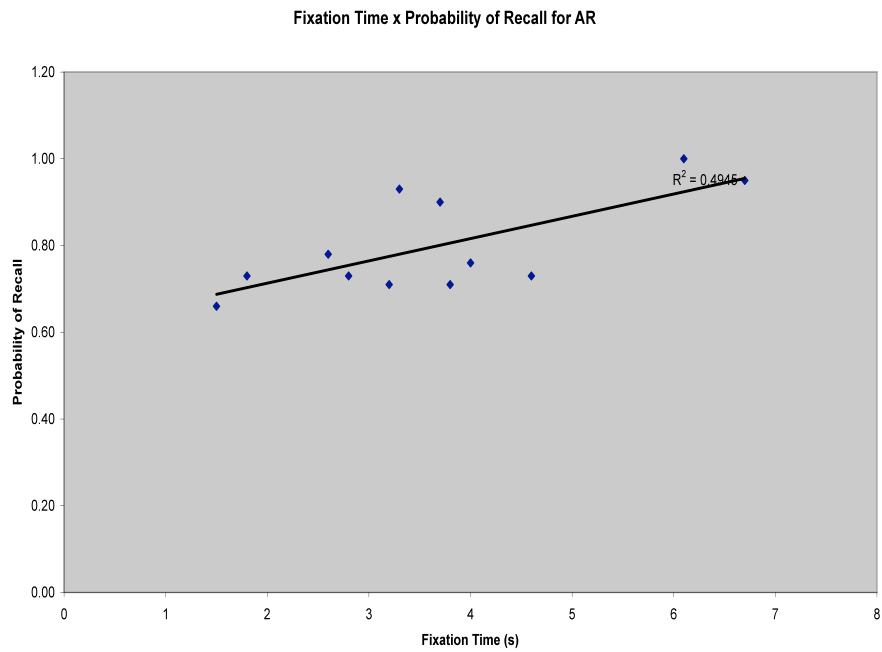


FIGURE 6.10: Augmented Reality Targets: Fixation Time versus Probability of Recall.

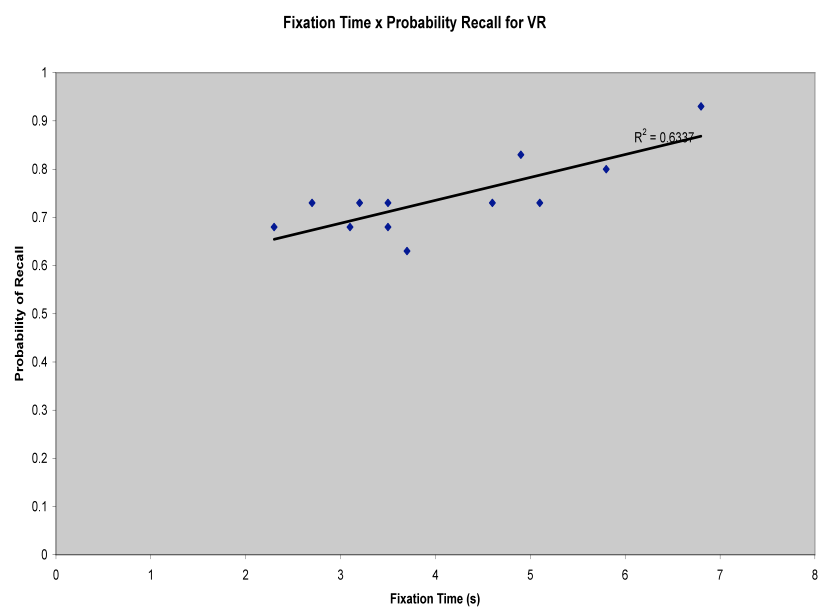


FIGURE 6.11: Virtual Reality Targets: Fixation Time versus Probability of Recall.

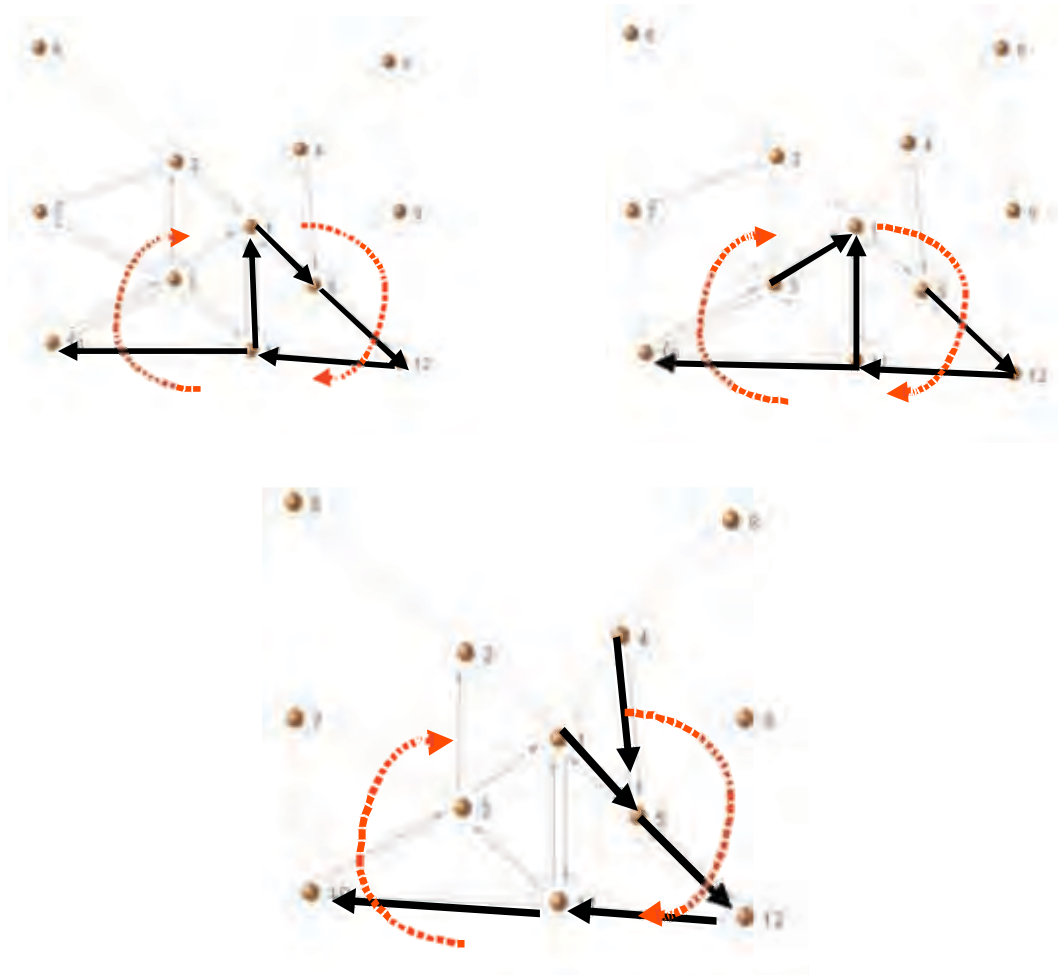


FIGURE 6.12: Scan Paths: Top Left Real Life, Top Right Virtual Reality and Bottom Augmented Reality, Note: Orange lines highlight clockwise motion as most popular search path.

(see Figure 6.12). There was a preference for central items over peripheral items and in the pattern of rehearsal was clockwise around the central and bottom items. The number of moves from one target to another were collected and tabulated (see Experiment 2, Scan Path Analysis Data appendix I, tables I.1, I.2, I.3). This formed a link analysis diagram that could be analysed as a network. These data were then entered into the Agna Social Network Analysis tool [143] and used to calculate the status of each node in the network. In this analysis, status refers to the number of connections to and from a given node relative to the total number of connections in the network. Average Number of Times that each marker is visited per test run read from column to row ie column position 1 goes to 2 once, 3 three times, four once etc.

Referring to the Scan Path Analysis data, see appendix I. In the AR condition it is clear that four nodes score higher than the other nodes, i.e., 1, 3, 5, 11. This group of highly connected nodes is followed by a second group, i.e., 2, 4, 10, 12 and 7, and there is a further group of nodes which have a low score, i.e., 9, 8, and 6. Focusing then on the VR condition, it is clear that four nodes score higher than the other nodes, i.e., 1, 11, 5, 3. This is similar to the AR condition, although the ordering is slightly different. This group of highly connected nodes is followed by a second group, i.e., 10, 2, 12, 4, and 7, and there is a further group of nodes which have a low score, i.e., 9, 8, and 6. Finally from analysis of the real life condition data, it is clear that four nodes score higher than the other nodes, i.e., 1, 11, 5, 3. This group of highly connected nodes is followed by a second group, i.e., 4, 12, 10 and 2, and there is a further group of nodes which have a low score, i.e., 7, 9, 8, and 6. This pattern is somewhat different from the AR and VR conditions in that node 7 now lies in the low scoring group.

6.9 Discussion

The strong preference for a clockwise pattern may relate the occlusion of the left eye. Wearing a HMD affects scanning of the real world [144], this affects scanning behaviour in the sense that more head movement is required to compensate for the occluded field of view of a blind spot. At first these data would appear to contradict the dwell time data. One interpretation is that although participants chose the same strategy for inspecting the items, this manifested itself differently in terms of dwell time suggesting that the optics interfered with the execution of their strategy. One caveat to the data is that given the musculoskeletal demands imposed on moving when wearing a head-mounted display, the extension of the neck is uncomfortable so participants tend to fixate targets below the horizon using head movements and to fixate on targets above the horizon using eye movements. Due to the added weight the HMD imposes on the head it is likely, participants tended to fixate targets below the horizon using head movements and to try to fixate targets above the horizon using eye movements; this is a natural pattern of physical activity when scanning a vertical surface and explains the difference in movement when wearing or not wearing head-mounted displays [145].

The analysis show that contrary to the null hypothesis suggested there was in fact variance between the applied conditions. Referring to hypothesis b), the tracking data

shows that search strategy did vary slightly between conditions, one observation is the fixation time disparities between experiments. This disparity was most likely caused by a subtle difference in the approach strategy in each condition. One cause of this is the fixation requirement of the markers in the AR condition, (stable peripheral view of bottles available). The marker based approach required a fixate then sample approach towards the objects whereas the real life and VR approaches allowed for a scan whilst moving strategy. The analysis of the scan path data also revealed that in all three conditions, the set of nodes 1, 3, 5, 11 have the highest status, which indicates that they are the most highly connected nodes in the network. For the scan path data this indicates that these nodes are visited more times than other nodes, and therefore serve as home points for the scan path, i.e., these are nodes from which the scan moves to other nodes and back. The set of nodes 6, 8, 9 score lowest on this analysis. Of particular interest in this study is the post-hoc analysis which showed that the VR condition resulted in somewhat different activity. One suggestion is that the AR and RL conditions encouraged people to scan the environment relative to their own location, i.e., ego-centrally. This used the position of the person in the world to scale judgments and decisions concerning the relative location of objects. In the VR condition, the environment might have been perceived relative to itself (even though the head tracking and positioning of displayed information was identical in all three conditions). This would suggest that the participants scanned the environment exo-centrally, in terms of the relationship between the virtual objects to each other. While this implies a subtle distinction, it is important because it suggests that the virtual environment is seen as a space that moves around the person, whereas the real and augmented environments move relative to the person. In this respect, body-based movement becomes something which is seen as effective and appropriate in the AR and RL conditions but inappropriate in the VR condition. In a similar manner, experiments reported in Boud et al. [146] indicated that using the tracking of a real object in physical space to control a virtual object resulted in quite difference performance to control using a mouse to move the virtual object; in this study and the current study, a plausible explanation for these differences arises from the participants perception of the space around them.

The apparent similarity in performance in AR/RL supports an argument that AR is more suitable than VR for training. In the AR/RL domain we are somewhat expectant and comfortable with environment and as such we are acquiesce to the scenario presented

to us in this experiment and therefore, adopt normal human behaviour. Seeing the world in VR however, is rather abstract and unusual and as such may intrinsically promote a different response from the user. This idea that the VR scenario feels more alien to the user may incur a tendency to trust the environment less and therefore interpret it differently. Interestingly mean VR performance shows longest overall dwell time (see figure 6.8) and poorest overall recall performance for VR, AR, RL respectively (refer to table 6.3). This analysis of results suggests that users approached the task in the same way as RL when utilising AR but probably not when utilising VR. This finding supports hypothesis a) in regards to transfer of training for AR-RL but possibly not for VR-RL. Referring to the initial hypothesis that where the requirement of the technology is closely aligned with the requirement of the user various forms of MR can be used to provide effective training platforms. This experiment shows that the way in which the user interprets the environment also affects human performance. An interesting observation here is to ask; in a fully realised AR implementation that supports full body movement how closely is AR/RL performance aligned. A further test of AR then is to see how performance is affected in fully mobile real world tasks.

Chapter 7

AR Performance Study - Search of a Local Environment

7.1 Introduction

Experiment one has shown that, the spatial visual cues that are facilitated by movement are a determining factor for recall. Scientists have shown that if an image is perfectly stabilised on the retina, the image disappears, therefore human perception requires that images be dynamic [147]. Human vision can only detect change and can only see something when that something is moving (either the object and/or the eye). Findings in earlier experiments proved that the addition of a real background image aided human performance only when the image moved, the point being that in-fact movement is the principle factor that affects the way humans perceive and perform in the real world. The governing finding therefore of the work done so far is that it is not the synthetic augmentation to visual perception that affects performance but rather the way in which movement cues are employed and adapted. This was shown in a study by Ruddle and Lessels [75] where performance in photo-realistic virtual environment (VE) was compared to an impoverished equivalent that contained far less visual information. The task involved completing a complex spatial search task, where a number of items had to be located amidst an array of decoys. Results showed no difference in performance in the photorealistic vs impoverished environments. However, comparison between a condition with no body-based movement showed a significant drop in performance. This

proved that body based movement, contributed to human performance significantly over the visual resolve of the virtual interface. In another study by Lemos et al. [141] they showed that photo-realistic images were associated with a higher workload for image identification tasks and additional visual detail did not increase performance for either identification or performance tasks.

The findings in the experiment 2 deduced that where technology was closely aligned with user requirement, AR and for the most part VR were successful. However, further investigation suggested that performing a simple search task in real, augmented or virtual environments involved participants attempting to follow similar search strategies resulting in similar recall performance, but with some differences in movement (particularly between VR and the other conditions). In this study participants were required to move around the environment and the focus was on RL and AR such to explore the possibility of differences between these conditions. The element of movement would be consistent between the two conditions. Perception of the environment was manipulated solely on the visual platform, real objects were replaced with synthetic equivalents presented upon correct registration of fiducial markers.

The trial in this experiment examines human performance in a small local environment facilitating a full range of motion, the intent being that AR could be suitable for real world training scenarios if performance is closely aligned with real life. Two conditions are presented, real life (RL) and augmented reality (AR). To ensure active visual search within the test environment an emphasis was placed not only on spatial location but also some specific object characteristic .

This section aims to answer the following research question:

If AR can be utilised such that the implementation of the technology is well aligned with that of the user, is human performance affected?

7.2 Participants

Twelve subjects each took part in a repeated measures study of three trials per condition. All participants were graduates aged between 21 and 26 years of age with an average age of 23. Of the twelve participants 6 were male, 6 were female.

7.3 Hardware Design and Software Development

Head Mounted Display *Lite LE-700A* The Liteye LE-700A was utilised (refer to chapter 6) as the display interface and was deemed suitable for a fully mobile experiment because of its light weight and its convenient power over USB capability.

Wearable Computer X^4 (*Chi-Four*) This experiment required complete freedom of motion to facilitate this a fully wearable system was implemented and a wearable computer was utilised. This was an in-house design developed by the HIT team at The University of Birmingham. This custom wearable PC combines a unique small form factor with powerful x86 based hardware architecture using a PC104 main board (see figure 7.1). Ports include SVGA out, two serial ports, on-board LAN and four USB ports. The 32-bit 1 Gig Pentium M-Class processor runs Windows XP and can therefore run any windows based software. The hardware set-up proved to be sufficient to run the AR fiducial recognition system developed earlier.



FIGURE 7.1: The X^4 (*Chi-Four*) wearable PC developed by the HITLab at The University Of Birmingham.

Head Tracking *Intersense InterTrax²*

The Intertrax² is a three degree of freedom device that can measure pitch, roll and yaw over 80° , 90° and 180° respectively. The device has a 3° angular rate and $256Hz$ internal rate with a latency of 4ms. The small form-factor of $9.4 \times 2.7 \times 2.7cm$ makes mounting to a HMD relatively easy and weighing only $29grams$, added head strain the overall system design is circumvented. Unlike the Wiimote used earlier the device uses

a USB communication bus through which it is also powered and therefore cannot be used wirelessly. For the purposes of this experiment however, this was not an issue. The Intersense SDK provided *C++* libraries. This was used in the .NET framework by developing a managed wrapper for the *C++* driver library. The readout from the device was then pooled and recorded to a text file.

7.4 Stimulus Objects

In this experiment the use of AR in the search of a small local environment was examined. To conserve objectivity in analysis between this experiment and the last the same twelve coloured bottles were utilised. To ensure active visual search within the test environment an emphasis was placed not only on spatial location but also some specific object characteristic, this was achieved through counterbalanced marking of the bottles with either an X or O. Participants were asked to search an office space for a set of six coloured bottles, the task was alternated between finding six items marked with either an X or O. These bottles were concealed within the environment so they could not be easily found without active search (e.g., looking under tables, behind chairs, on bookshelves and so on). Items of furniture were placed in the room to present obstacles that ensured participants followed a similar path through the scene.

7.5 Design

When searching for evidence there are various search methods employed by crime scene investigators. The search techniques employed depend heavily on the size and shape of the area and also the nature of the crime (see figure 7.2).

- Spiral Method - Involves walking from the centre of a crime scene spiralling outwards. Only really suitable for small areas as evidence can be overlooked as circle widens.
- Strip Method - Effective in large open spaces, investigators walk along the length of the crime scene then back along parallel strips until the area is covered. Also often applied in smaller areas with one investigator.

- Wheel Method - Investigators begin in the centre of the scene and gradually work their way outwards. Shortcomings are that evidence could be destroyed in the densely covered central area and evidence could be overlooked towards the peripheral zones.
- Grid Method - The best and most thorough search method for large areas with multiple officers as the whole scene is canvassed twice. Investigators walk in parallel strips along one axis then again at right angles to their original route.
- Zone Method - Divide crime scene into zones or sectors, good for large indoor locations
- Line Method - Used mainly for outdoor locations, investigators walk abreast in a straight line until the area is covered.

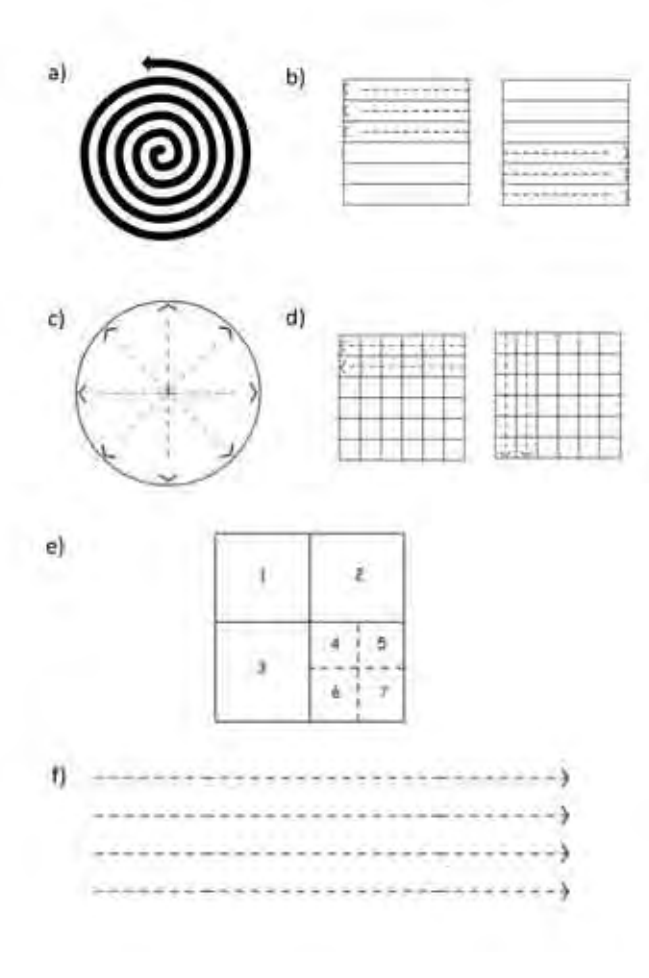


FIGURE 7.2: Search Methods: a) Spiral b) Strip c) Wheel d) Grid e) Zone f) Line

One of the main scenarios envisaged for augmented reality is where systems are used by individuals moving around their environment. In this experiment the use of AR in the search of a small local environment was examined. Participants were asked to search an office space for a set of six coloured bottles. The task was alternated between finding six items marked with either an X or O (see figure 7.3). These bottles were concealed within the environment so they could not be easily found without active search (e.g., looking under tables, behind chairs, on shelves and so on, see figure 7.4).



FIGURE 7.3: Representation of bottle marked with an X and O respectively.



FIGURE 7.4: Synthetic bottle as represented on a shelf examined in trial scene.

Based on the scene the evidence was positioned within it in such a way that participants would emulate the strip method of search. This was achieved via subtle obstacles that ensured participants followed a linear path through the scene. Participants were asked to locate every bottle marked with an X or every bottle marked with a O. They were not informed that in each case there was always a total of 6 circled bottles and 6 crossed bottles. This ensured that a user would cover the room fully before returning to recall findings. There was no time restraint placed on the user and no guidance given as to how to search the environment. Each user began the experiment from the same position,

facing marker 1. The layout of the room was such that a route around the environment was observed (see figure 7.5), users covered the floor space in a linear manner, such to replicate the chosen search method observed in scene of crime investigation. Each experimental configuration comprised twelve objects in total of which six were marked with an X and the other six marked by O. Each bottle was of one of three primary colours, this colouring scheme was distributed evenly, so for example of the four red bottles, there were two red X bottles and two red O bottles. Upon successful location of all six objects, participants were asked to report from memory the location of the six X/O objects (see Recall Sheet - Experiment 2, appendix J). Correct location and colour was necessary for a right answer.

Repetition was kept to a minimum in this experiment. Configurations in the real environment required that objects to be moved around to correspond to each pre-defined layout. This placed a restriction on the logistic practicality for the number of repetitions that could be carried out. Also since the participant was actively searching the environment, fewer repetitions ensured that factors such as fatigue did not compromise any of the results.

7.6 Hypotheses

In order to test the research statement stated earlier the following hypotheses were formulated.

a) Previous work, reviewed earlier, would suggest that where the technology is closely aligned with the requirement of the user, performance between the AR and RL conditions should be indifferent since human performance should be governed by body-based information.

7.7 Procedure

Each participant was briefed on the procedure of the experiment. They were informed that they could withdraw at any time or could refuse permission for their results to be used, following the explanation, participants signed a consent form (see appendix E).

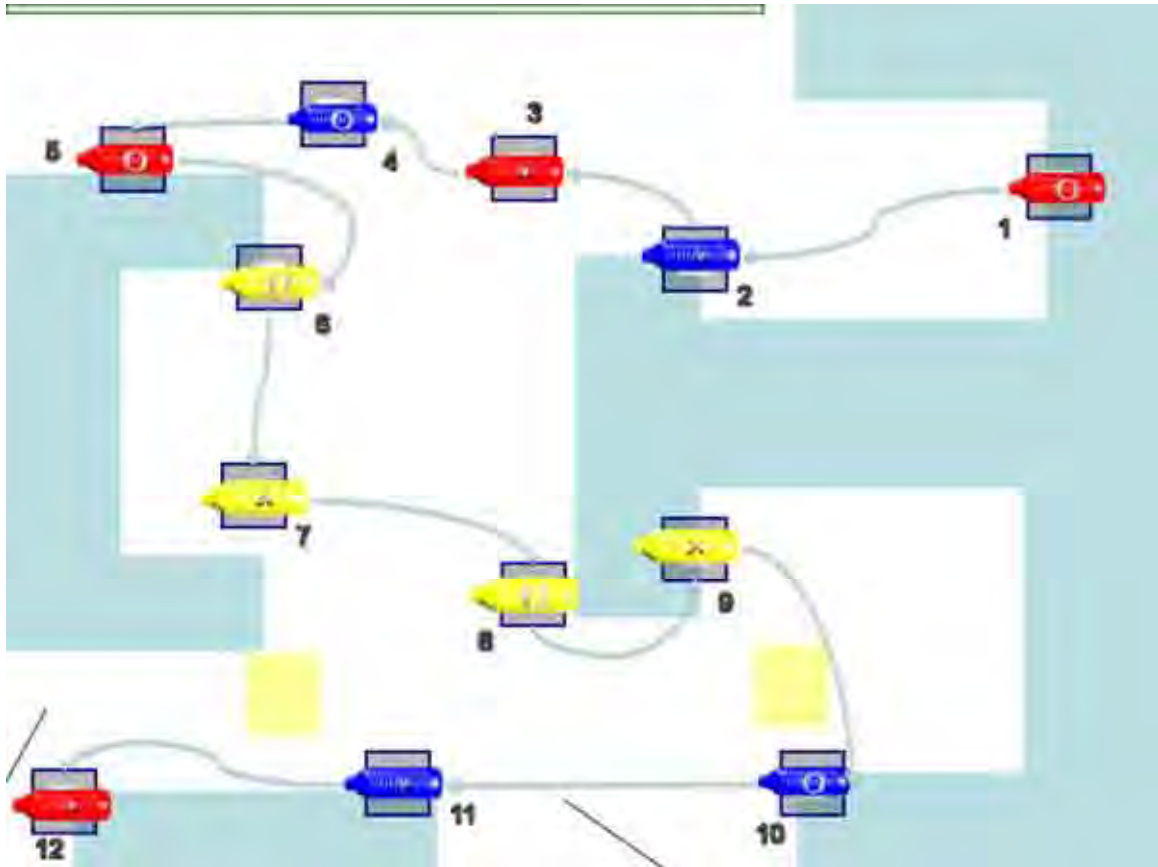


FIGURE 7.5: 3D Virtual Bottles Augmented onto Markers - Furniture arranged such that a linear path through was ensured and the complete room canvased on each trail run.

Users wore a HMD device affixed with a webcam and the Intertrax head tracker (see figure 7.6). These devices were connected to the X^4 wearable computer that was housed in a specially designed vest, worn on the users back. A short introduction session was conducted to make sure that the kit fitted comfortably, that the HMD was viewable and a short familiarisation session was conducted to demonstrate how the fiducial markers worked.



FIGURE 7.6: User Trial - X^4 wearable computer fitted into vest for non-invasive mobility.

Markers were laid out throughout the test environment in a set pattern that was preserved between conditions and repetitions of the experiment, the layout of the room was such that a route around the environment would be followed (see figure 7.5). For each condition three configurations of bottles was arranged, for each user a set alternating procedure was followed, each participant began with a different condition, configuration and search. Bottle configurations in AR and RL were not the same and a participant would not be presented with the same bottle configuration more than once, ie. the following example would be avoided eg. Test run 1 AR - Config 1 - find all X marked bottles, Test run 2 AR - Config 1 - find all O marked bottles. Each participant began

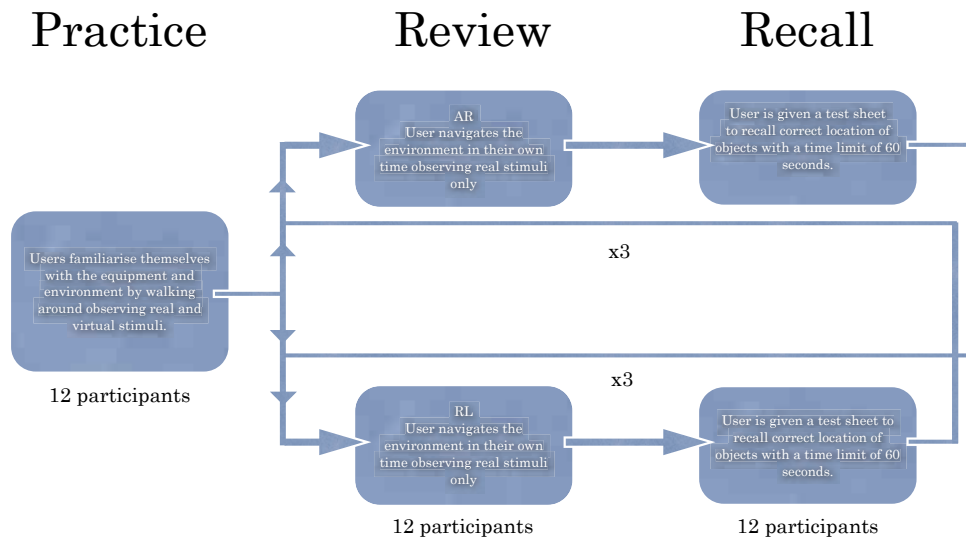


FIGURE 7.7: Design of study showing how the tasks were carried out with each participant. The testing sequences of the AR and RL conditions were counterbalanced between participants

the experiment from the same position, facing marker 1. Once all 6 marked objects had been successfully located, users were presented with a test sheet and asked to record from memory the location and colour of the six 'X' marked or 'O' marked bottles.

7.8 Results

7.8.1 Measurements

An initial run in each condition were assigned to the practice session and used for familiarisation purposes only. Thus, those results were discarded from analysis. A total of 72 sets of data were collected from the experiment (12 subjects x 2 methods x 3 trials) and hence, 36 data sets for each condition. The maximum score for a single data set was six, any ambiguity in a recording would be regarded as incorrect recall, such as 2 answers in the same box.

7.8.2 Analysis

The data, averaged for each participant was put into two paired t-tests (one for correct response and one for time). There were no significant differences in the time taken to complete the task (see figure 7.8). This was confirmed by a t-test [$t(11) = -0.312, p = 0.7609$]. Although recall performance on the whole was poor with a mean average recall out of six items of 2.75 (46%) for AR and 3.52 (59%) for RL conditions, there was a significant difference in recall performance of the real and synthetic bottles, with real bottles being more likely to be recalled than their synthetic equivalent. This is confirmed by a t-test, [$t(11) = 2.775, p = 0.0181$].

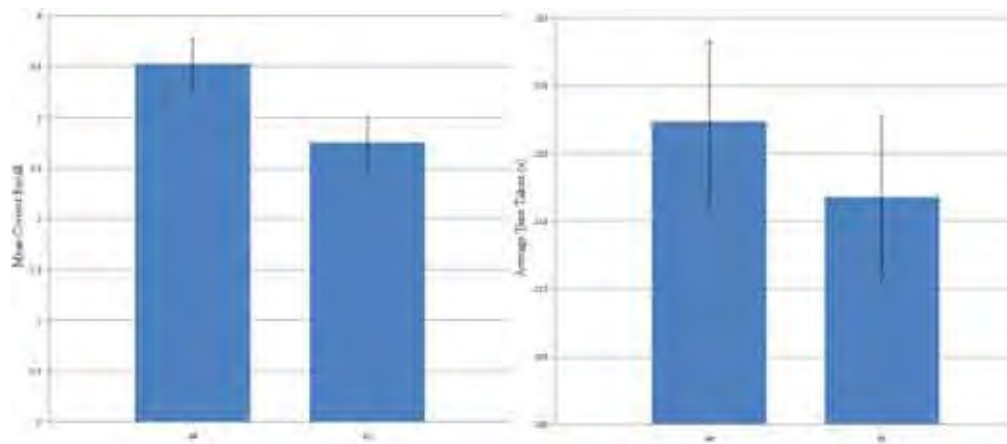


FIGURE 7.8: Left - Recall of Concealed Items - Recall Rates $t(11) = 2.775, p < 0.05$ (standard error, SEM of 1SD) Right - Search time (in seconds) to find all concealed items, $t(11) = -0.312, ns$ (standard error, SEM of 1SD)

The registration requirements of the AR system meant participants needed to 'line up' the head-mounted camera with the markers, the tracking data (see Pitch and Yaw tracking data, figures 7.9, 7.10) shows that in the AR condition head movements were restricted in magnitude. This may have had two implications: first, that participants spent more time searching for fiducial markers than viewing the synthetic bottle stimuli that were to be remembered and second, because head movements were limited once the marker was registered by the system and the synthetic stimulus presented, participants were not able to view the stimulus from different angles restricting the amount of contextual information bound to the stimulus to aid its memorisation through a process of elaboration (see experiment 1, chapter 5).

Yaw Tracking in RL and AR

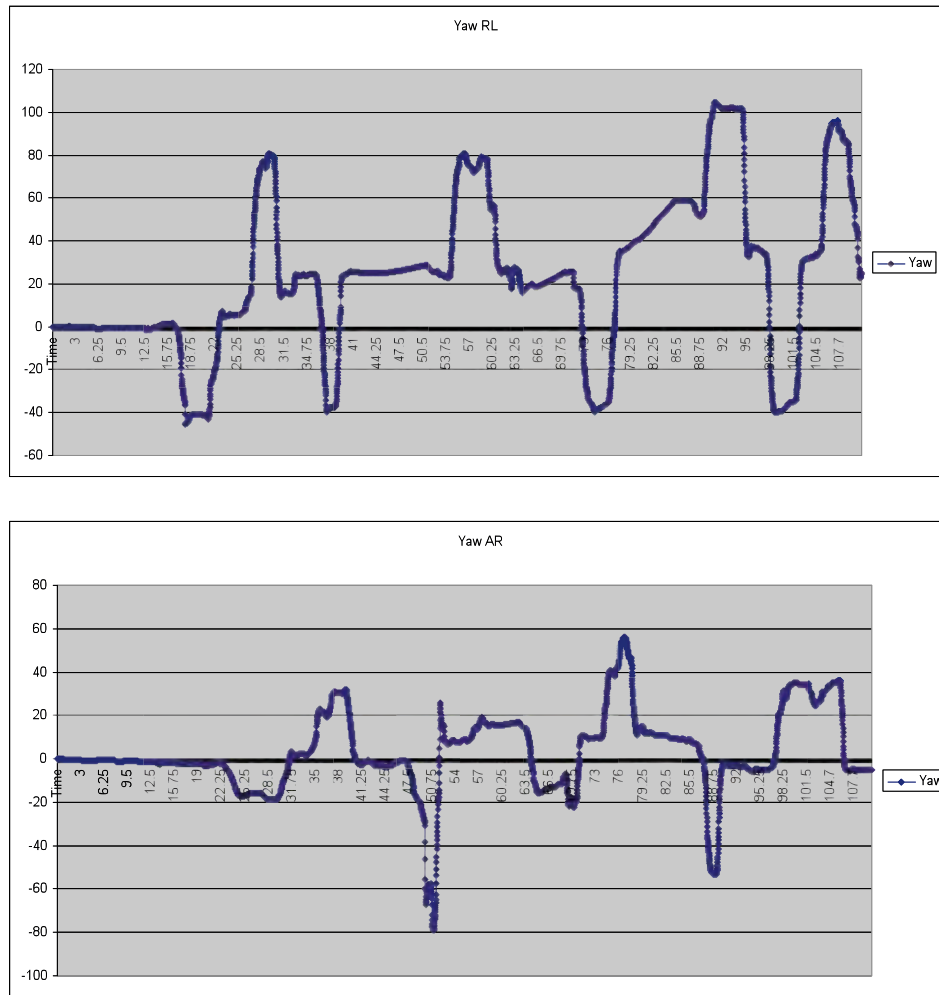


FIGURE 7.9: User Sample Head Movements Across Yaw Domain in seconds, Top Panel, Real Life Condition, Bottom Panel AR Condition

Pitch Tracking in RL and AR

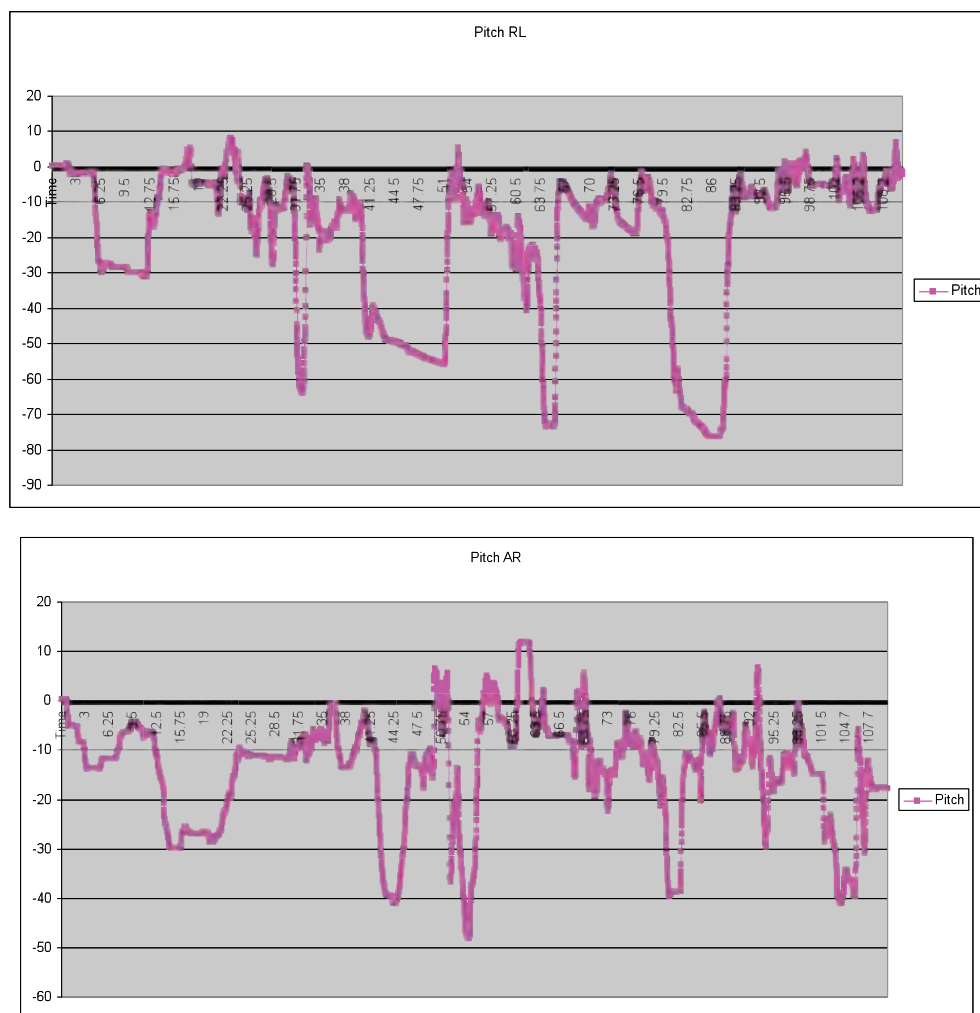


FIGURE 7.10: User Sample Head Movements Across Pitch Domain Over Time in seconds, Top panel, Real Life condition, Bottom panel AR condition

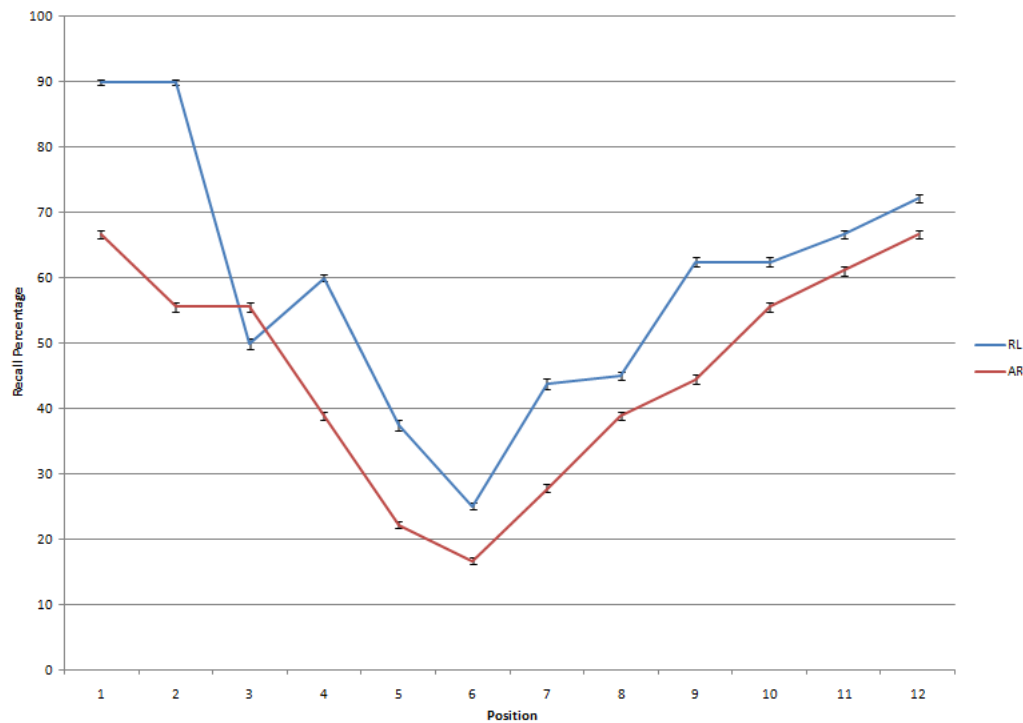


FIGURE 7.11: Recall Performance against Object Position, graph showing how recall differed between RL and AR conditions

The analysis of position data proved somewhat difficult as the serial position of items was not distributed evenly between condition configurations. However, since participants navigated the search area in a linear manner a serial position effect was still observed see figure 7.11.

7.9 Discussion

In this experiment the time spent completing the task, recall performance and head movements were recorded. The time spent searching the environment was similar, but recall performance differed. Two conditions were compared, one in which the bottles were real and one in which synthetic bottles were presented in AR on the basis of software recognising a fiducial marker in the scene. Thus, the presentation of synthetic items was subject to registration of markers by the AR system that required the marker to be fully visible and square with the camera (the angle of marker registration tolerance was measured to be approximately 40°). Often one could not therefore view markers in the visual periphery and although marginal, there was a delay between alignment and the

item appearing. Although in the design stage the affect this would have was thought to be nominal, the results insinuate this hypothesis was misconstrued. Recall performance in AR was poor compared with RL and one possible explanation of the difference lies in the head movements that participants performed during their search of the environment.

The head tracking data showed less movement in the AR condition, which is the result of the need to find a fiducial marker and stabilise direction of gaze in order to register a target. This placed a constraint on movement that was not apparent in the RL condition where participants tended to make larger head movements to scan the environment (although, of course, the movements were still constrained by FOV of the head-mounted display that they wore). One interpretation of this finding is that the fiducial markers, rather than being bearers of virtual information, took on two roles: one of object in the environment, that the participants looked for, and one of the surface to present targets. It might have been possible that participants were performing the search task for the marker-as-object and using this to close each step in the search; rather than responding to the targets. However, the distribution of recall results (figure 7.11) suggests that some other factors were at play in addition to this. If participants were not responding to the targets then one might expect performance to be around chance level throughout the trials, but the apparent serial position effect shown in figure 7.11 implies some memory problems, particularly in the objects located in the mid-range of the search routine. An alternative explanation was that the effort to detect markers and then ensure registration could have been sufficient to place a demand on the person such that earlier recall of targets was disrupted. While the results are not detailed enough to resolve this issue, the suggestion is that the combination of moving around an environment, searching for specific objects and then processing these in sufficient detail to enable subsequent recall, is more affected by the head-mounted display in the AR than in the RL condition.

Another possibility is that participants in the synthetic condition interpreted the task as merely one of finding the fiducial markers and saw the synthetic stimuli as incidental to this; the presentation of synthetic stimuli may merely have acted as a cue to closure indicating to the participant the marker had been found and logged. If we take the analogy of memorising items in a room, it may be easier to remember the items through walking around and viewing the items and their spatial relationship to each other and ones individual frame of reference from different angles than in the case of viewing the room and its contents through a letter box (i.e., a fixed viewpoint). What is clear is

that the failure to perform as well in the AR condition is down to the impediment the technology placed on body based search behaviour. The findings therefore further cement and backup previous work suggesting that body based behaviour is a governing factor in rating human performance. From a visual aspect synthetic bottles could be viewed just as easily as real bottles, as the results showed, successful observation of all bottles was completed in a similar time frame. The AR condition caused users to pay less attention to the surrounding environment and therefore subtle elements such as peripheral recognition of objects and spatial registration of the surrounding environment was lost. This may also be supported by the finding that a serial position effect seemed prevalent in the AR condition and somewhat less prevalent in the RL condition. Thus, suggesting that since AR objects were interpreted with less context than their real world counterparts a serial position effect dominated the memorisation process.

The results do not directly support the hypothesis that performance between AR and RL recall should have been indifferent. However, this apparent failure of AR in terms of this experiment provides further understanding of a more salient global observation regarding human performance. Put simply human performance is affected not merely via the medium through which the world is perceived but more over, the constraints governing how movement in the world is controlled. Taking this further, with reference to the findings and deductions made about transfer of training in chapter 1, it is clear that the AR condition although in theory appears to be very closely aligned with the real condition, human behaviour is in fact quite different. A simple contention to this argument might be that it is a fault of the technology ie. the small delay in marker registration. This however would have no bearing on the peripheral failure in any registration based AR method to register markers and display virtual objects. Therefore registration based AR (marker based or otherwise) is unsuitable as a measure of training when full body based movement is supported because the requirement of the technology does not support normal human behaviour. Upon relating these findings to the results from experiment 1 these findings are not all that surprising, it is fair to comment that in the RL condition a spatial co-ordinate reference frame was supported (ie. *"In front of me I can see the red bottle, to the right of me is the blue bottle"*). While the AR condition did not support one, only the bottle being lined up in the HMD view frame could be seen and this affected the way in which the task was approached. No co-ordinate reference frame resulted in users essentially navigating the environment, blindsighted,

moving from one marker to the next with no regard for their spatial positioning in the environment.

Chapter 8

Conclusions

8.1 Satisfaction of Problem Statement

The problem statement asked firstly:

Can emerging technology such as AR could be utilised effectively to aid problem solving methodology such as crime scene investigation?

In effect the research answered this question by satisfying the second statement.

What affects do the properties of AR technology have on human performance?

The research deduced that the property of movement when supplemented with a real world frame of reference was key to aiding task based performance. It was found in experiment 2 that the variance in performance between AR and equivalent conditions such as real life and VR was relatively small but only if the requirement of the technology and that of the user was very closely aligned. We showed progressively through analysis of trial data that movement was indeed key to task based performance and memory recall.

Overall, the studies imply that people attempt to follow similar search strategies across the various conditions and so one might conclude that each would be equally effective as a training environment. However, the studies all reveal differences in relative forms of performance that suggest that the constraints imposed by these technologies can hinder the direct application of everyday skills and movements. By requiring all participants to

view the environment through head-mounted displays the studies place similar (somewhat unusual) limitations on participants ability to view the world. This allows the studies to compare the impact of different media on behavior in the environments. The main focus of the studies has been the definition of behavior in terms of movement, primarily in terms of head movements performed during search tasks.

Experiment 1A was concerned with the orientation element of movement and also the effect that a real world frame of reference had on human performance. Analysis of data showed that where a real frame of reference was unavailable the orientation component of movement did not aid performance at all. However, when objects were observed when a real world view was available, recall improved. This idea that movement coupled with a real world view improved performance was further tested in experiment 1B that blocked sets of objects together in groups. This experiment proved that a spatial frame of reference provided by a real world background improved human performance. An interesting corollary found in experiment 1A came from the results of the static condition which showed that the removal of movement had an effect on the way in which the task was handled by the user. This was attributed to a dramatic variance in recall primacy and recency items. This deduction regarding the importance of movement defined the goals of the experiment 2 which provided the user with an increased degree of freedom in movement but kept the alignment of technology and user requirement between conditions extremely close.

Experiment 1A and 1B deduced that VR could also facilitate movement and provide a synthetic moving background. Referring again to the background research VR has been used to good effect in scenarios where the technology and user requirement are closely aligned hence, the introduction of a third VR condition. In the experiment 2 the comparison of performance across the platform of the mixed reality continuum (Real Environment – > Augmented Reality – > Virtual Environment) helped to re-enforce the findings of previous work in the VR domain and highlighted that this idea of the importance of close alignment between technology and user requirement also held true for AR but also, suggested that a users interpretation of the world was also a factor. Analysis of both user scan path and dwell times concluded that where various media of MR were closely aligned, MR was a suitable platform to support human behaviour. However, this data also determined that human behaviour was affected by the constraints imposed by the technology. The tracking data showed that the weight of the apparatus encouraged

users to favour targets below the horizon and the occlusion of the left eye caused users to favour navigating the objects in a clockwise direction. Also a behavioural change was exposed by the variation in dwell time between the VR and other conditions implying that their activity was more likely to be more natural in the AR and RL conditions. In other words, movement seemed to be ego-centric in the AR and RL conditions, with movement being made relative to the person's location in the environment. However, in the VR condition, it is suggested that movement is more likely to be exo-centric and made relative to the space defined by the virtual environment. The difference between RL and VR, in terms of ego-centric and exo-centric perceptions of movement in space has been shown in several of the studies reviewed in the introductory sections. This study adds further support to the claim that body-based movement in VR is not the same as in RL. In contrast, for the movements in the X and Y plane, there seems little difference between AR and RL. This would suggest that AR could be an appropriate substitute for RL in training.

Unfortunately, when participants used movement in the X, Y and Z planes to explore an environment (experiment 3), there were strong differences in performance. Again, these differences can be explained in terms of movement. Experiment 3 was designed to emulate a training scenario akin to a crime scene investigation in which a user would navigate a real world scene that consisted of objects positioned around an environment that required investigation. In this experiment a fully mobile search and recall task supporting full body based movement was compared using AR and a seemingly identical real life condition. The only difference being that in the real condition an actual bottle was placed on the marker. This experiment determined that, as was found in the first experiment, even a very subtle difference between conditions can impact on human behaviour. In the RL condition, participants made larger head movements which indicate broader sweeps on the environment to look for bottles and shorter fixations on the bottles to recognise them. In the AR condition, in contrast, head movements tended to be smaller and to exhibit longer fixations. While this is partly a function of the registration required by the fiducial markers, it was felt that this was not sufficient to explain all of the variation in performance. Rather, it is proposed that the search task becomes defined slightly differently in the AR condition: such that participants are performing additional search sub-tasks to identify the target and these sub-tasks combine sufficiently to interfere with the simple memory tasks required.

In terms of body-based interaction, the studies reported in this work require movement by the participants in order to find objects (as opposed to using body-movements to control devices). While this represents a fairly basic form of interaction, it does provide a useful platform for comparing performance of participants when dealing with media presented in the form of real (video), augmented or virtual forms. Given that the key differences in conditions in these studies lies in the nature of the media used, it is interesting that significant differences have been discovered. The main conclusion from these studies is that people are likely to attempt to respond to large spaces that allow movement through them in a similar manner, e.g., by following similar search strategies, but that the variation in media can lead to subtle changes in behaviour that affect performance. For this reason, it is important to consider how equivalence can be made between the real environment and the augmented or virtual counterpart for training purposes. If the environment can be interacted with from a seated position, then movement can be performed in a sufficient similar manner in RL and AR, although it appears to differ in VR. This has been shown to be the case when comparing RL and VR and by the differences between RL and AR. To some extent, this arises from controlling of potential confounding factors that influence movement. When the person is required to walk around the environment, then movement becomes affected by the media of presentation and the suitability of AR diminishes.

Taken together the work in this thesis suggest that outcome measures of performance can, in some circumstances be similar across real, virtual and augmented environments. However, the manner by which these outcomes were achieved can vary. If one is intending to train physical skills, then practice in synthetic environments can lead to subtle changes in behavior even when the person is allowed to employ body-based movement. This could, potentially, lead to problems with transfer of training from the synthetic to the real environments.

As technology improves so will the ability of that technology increase such, that the user requirement will be met more satisfactorily. Referring to the first research question then, it appears that AR could be used effectively to aid problem solving methodology however, the alignment between the technology and task requirement needs to be carefully considered if the media is expected to be used to good effect. The research has also concluded that it is the ability to move through the environment, a feature inherent to AR technology that best facilitates human performance, in search and recall tasks.

Therefore, as long as technology can be suited to the user such that their requirements are closely aligned AR can be used as a tool to aid problem solving methodology such as crime scene investigation.

8.2 Further Work

It is possible that an AR system could be designed that facilitates full body based movement while satisfying the requirement of the user. The lynchpins so to speak of the system requirements would be governed by the findings in this research:

- The user requires a large enough field of view such that a real world reference frame is facilitated.
- Synthetic representations of objects must be manifested over the complete view frustum and not just in direct alignment with the users view direction.

A binocular see-through implementation would limit factors that affect normal human behaviour as this technology does not impede real world behaviour, the real world is perceived first hand and in stereo with a reasonable FOV. The NVIS' NVisor ST is a binocular see through HMD that would satisfy these requirements. However, findings in this research suggest that the HMD should be light enough to avoid a bias to natural head movement, this is a concern since high end HMDs such as the NVisor ST¹ weighs 1200g. However, with the popularity of AR modals increasing more recent developments present possible considerations with satisfactory specification such as Vuzix's Wrap 920AV² unveiled at the CES show in Las Vegas, January 2009 which provides binocular see through HMD technology and weighs in at around 3 ounces, is now available to buy from many major online retailers and also includes iPhone compatibility plus support for many popular PC titles (VR only). To satisfy the second requirement the environment could be prepared and scale modelled in a 3D engine and the user could be tracked by a wide area 6DOF tracker system such as the Hiball-3100³. The ability of this system to bring the requirement of the user back into close alignment would suggest that AR could be successfully implemented to aid task solving methodology.

¹<http://www.nvisinc.com/product2009.php?id=5>

²http://www.vuzix.com/iwear/products_wrap920av.html

³<http://www.3rdtech.com/HiBall.htm>

The main body of this work focused on utilising AR such to enhance it with virtual representations of real world objects, as such this is conceived as being the most salient and profound attribute governing most implementations of AR technology. However, AR also has the ability to enhance the world by appending real world objects with contextual information. An implementation of this kind could be applied using the system developed in chapter 4 that utilised a neural network to recognise real world objects. Rather than present synthetic representations of objects to the user it is envisaged that it would be beneficial to a crime scene investigators methodology if they can see a real world object enhanced with contextual information about a crime. Tests concluded that there were limitations and performance issues that need to be addressed before this system could be utilised to good effect. Therefore findings remain inconclusive in this area and require a more involved and timely development venture.

The final experiment explored and analysed in chapter 7 concluded that upon using time as a measure AR performed satisfactorily with all participants completing the task in approximately the same time. However, analysis of recall performance deduced that there was a significant failure of AR compared with RL in recall ability. This experiment showed that despite the apparent concordance between the AR and RL experimental conditions, the fiducial marker technology employed did not satisfy the requirement of the user. Thus, this suggests that a fiducial marker system (the de-facto standard implementation of AR) is unsuitable for transfer of training in fully mobile real world environments. Contrary to the success of the experiment which showed that even the slightest technical disturbance, can have a significant effect on a users approach to the search of a real environment and consequently their performance. The salient finding of this experiment is the apparent contradiction between the lack of effect of time but significant difference in recall. It was deduced that although minor, a delay between movement and registration of objects in the AR condition altered the perception of the task because users were more detached from the environment . Users took roughly the same time to search and discover all items however the encoding of items to memory in AR was compromised. The likely cause of poorer recall in AR was due to:

- Users were less aware of objects' spatial relationship to each other. The use of marker based technology meant that it was only possible to register the object attached to the fiducial marker close to and directly in-line with the user's view

direction. This created a letterbox environment where users were forced to encode objects in relation to each other as opposed to the environment. The effect of which was shown in experiment 1 where a real world frame of reference had a significant impact on performance (refer chapter 5).

- Marker based registration requires that users approach objects in a certain manner in order to achieve accurate registration. It is possible that this requirement had an effect on the behavioural approach to the task because users interpreted the task as one of finding the markers and saw the synthetic stimuli as incidental to this. Users in effect spent less time encoding item characteristics to memory because time had to be lent to lining up objects for registration. The effect of which was shown in experiment 2 where dwell time on an object had a significant effect on the likelihood of its correct recall (refer to chapter 6).

A particularly prudent point then is a designer wishing to explore the use of AR in fully mobile real world environments further, would need to employ methods that take into consideration what has been highlighted as a result of this research:

- A system that does not rely on pattern recognition to achieve registration would be preferable, such as the ultrasonic 3D positioning system explored in chapter 1. If the designer wishes to use a pattern recognition system a number of technological hurdles must be overcome in order to circumvent the caveat caused by often having to line up or approach targets in a certain way in order to achieve accurate registration. Refer to discussion section, experiment 3, chapter 7.
- A binocular HMD should be employed so that the environment may be observed without compromising search behaviour, this was highlighted in experiment 2 with the apparent failure of VR to support the same human behaviour (refer to chapter 6). This may be difficult to overcome as current binocular technology is flawed, the review of literature in chapter 3 gave reason to suggest that a binocular HMD implementation using currently available HMD technology is unsuitable for search of a real world environment. Factors such as a restricted field of view, visual fatigue, motion sickness, monoscopic and binocular visual field overlap and binocular rivalry would need to be taken into careful consideration before employing a binocular HMD.

- The real environment should be presented in its natural state and if possible should not be compromised or altered to facilitate the registration technology. It was shown in experiment 1, chapter 5 that whilst ostensibly irrelevant a real world frame of reference did aid performance. Thus, sanitising the environment to aid the requirement of the technology could have an adverse effect on the user's requirement to perceive it in a natural state.

8.3 Final Thoughts

The start of this thesis was concerned with the technology used in AR and the relative effects this technology has on visual perception and consequently human performance. An extensive review of literature determined that the constraints AR technology imposed on visual perception could have an effect on human performance. However, despite this the literature also showed that the most compelling attributor to task based performance and recall was movement. Since movement is a key property of AR technology the research attempted to examine this theory of movement in relation to spatial navigation more closely. Findings from the static condition in experiment 1 and the poor performance in the AR condition in experiment 3 show that even subtle differences affecting human perception of the world such as the absence of a moving real world frame of reference have a dramatic impact on human behaviour and consequently human performance. This leads us to the question of whether AR could be utilised to aid task solving methodology such as crime scene investigation.

The application of AR to assist the investigation of a scene of crime was explored in the experiment 3. It proved that the technology employed introduced a bias to the way in which the scene was explored in real life and AR experiment conditions. In conjunction with the findings from the earlier experiments, overall it proved what the research of literature suggested:

- In a scene visual perception of the world has a significant impact on recall.
- For the utility of AR to be sufficient the ability of the technology to satisfy the user requirement for the task is paramount.

- The limitations that current manifestations of AR impose to both freedom of movement and visual perception need to be addressed in order to govern whether it can be used to good effect, to aid task solving methodology.

In order for conclusive deductions pertaining to the utility of AR to be deemed appropriate for use in task solving methodology further research is required. This research would require that an advanced AR system be designed that was capable of meeting the requirement of a fully mobile user. The salient deduction of this research is that assuming the technology can satisfy the user requirement then it is reasonable to suggest that AR can be used effectively to aid task solving methodology.

Appendix A

Methodology Diagram for Forensic Investigation Improved Through the Advent of Ubiquitous Computing Methods

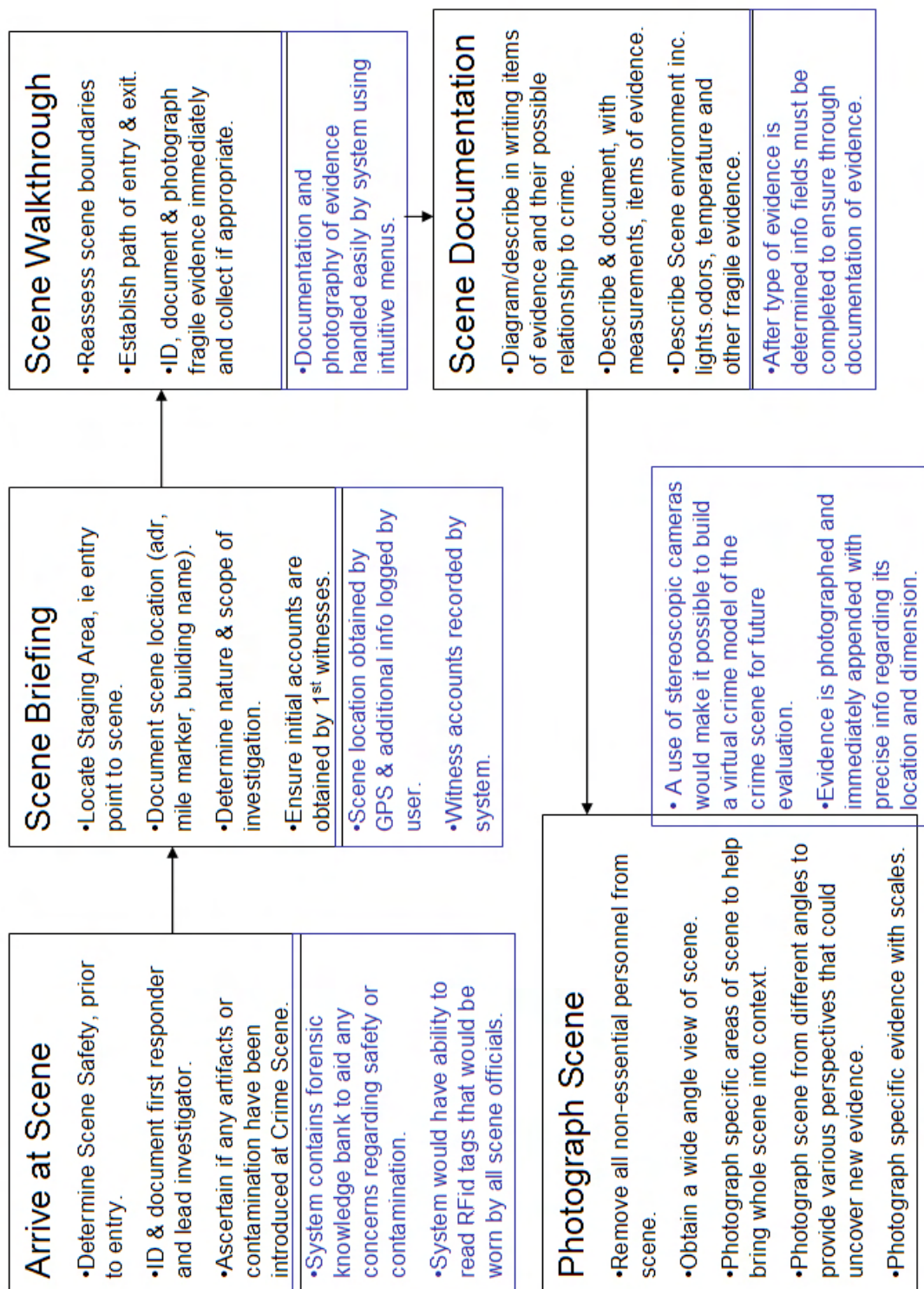
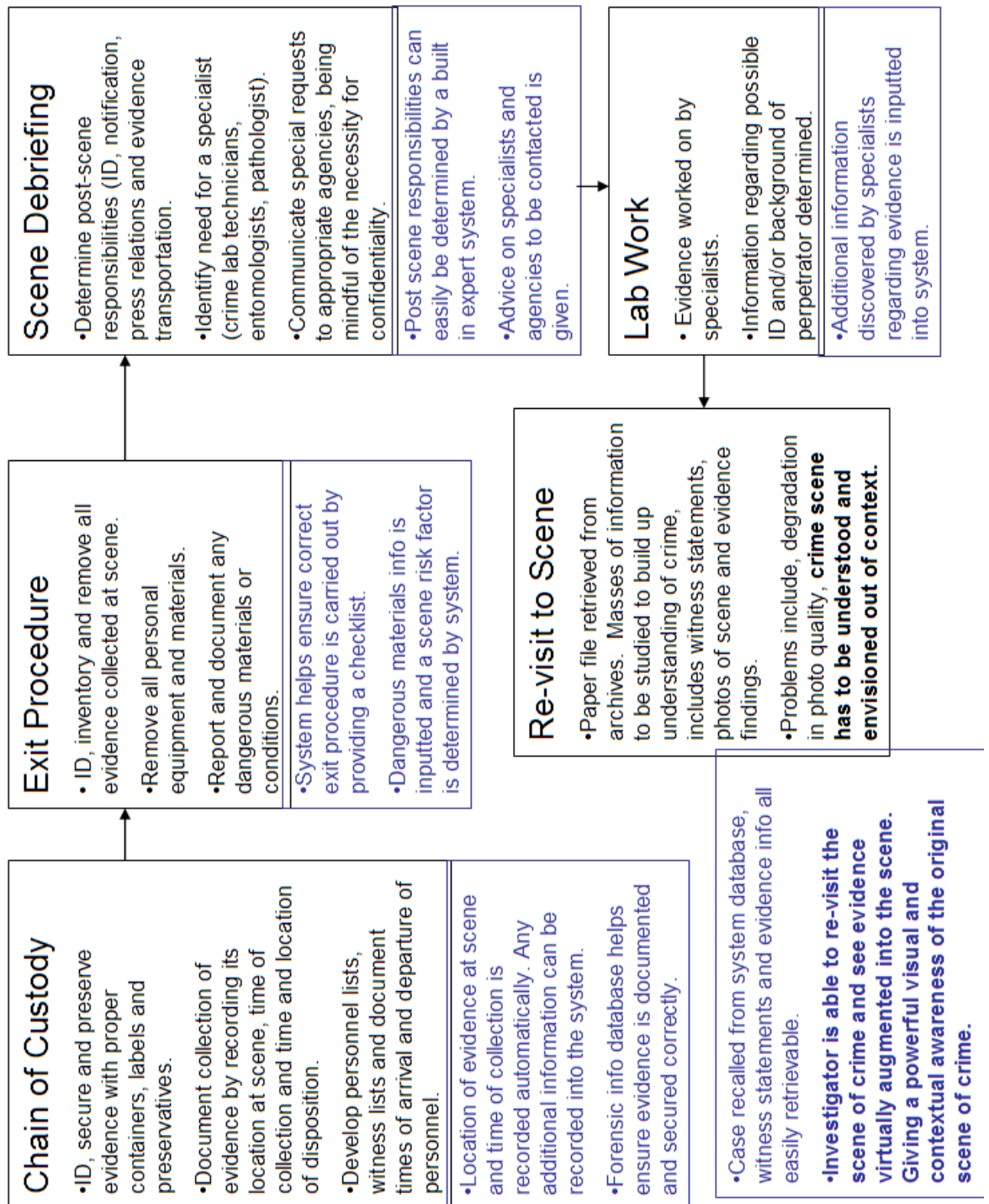


FIGURE A.1: Forensic Methodology Flow Diagram Improved through Ubiquitous Computing



Appendix B

Ultrasonic Transmitters

Positioned Around a Real World

Environment

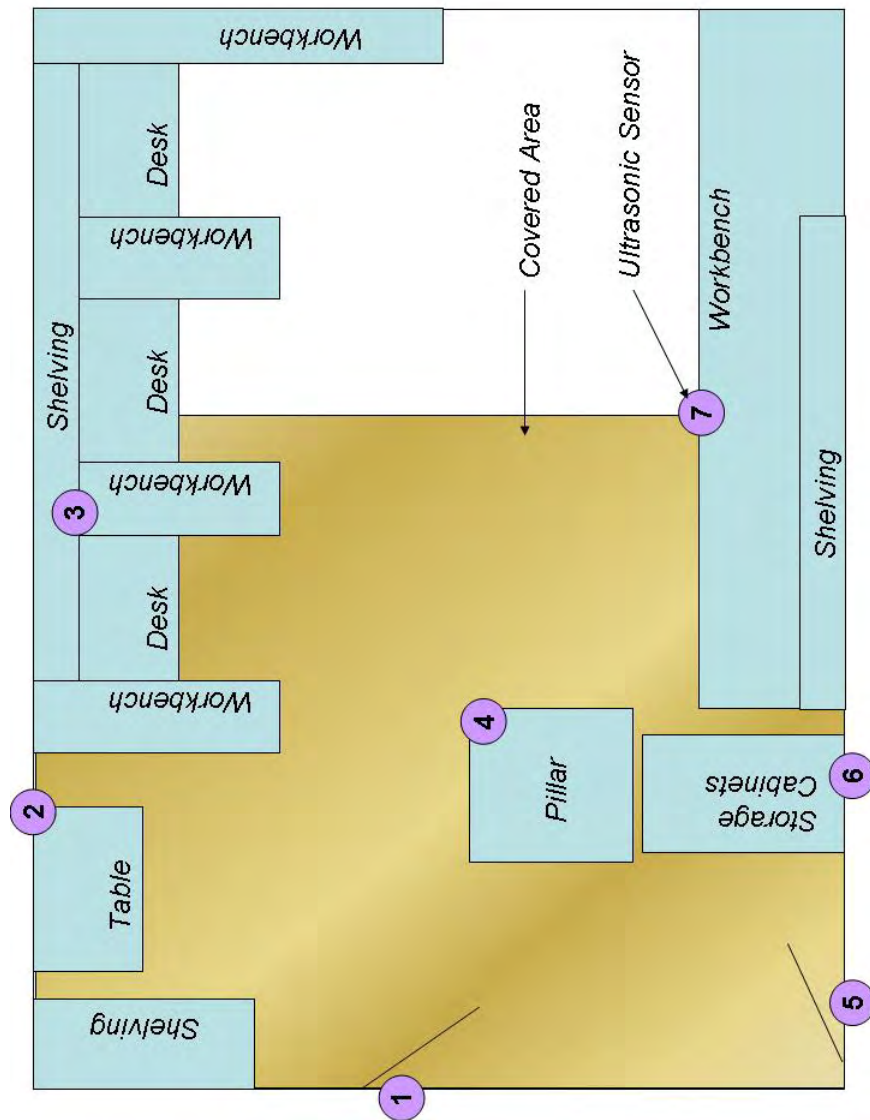


FIGURE B.1: Ultrasonic Transmitter Configuration

Appendix C

Neural Network Node

Configuration with Relative

Weights for Ziffer 1

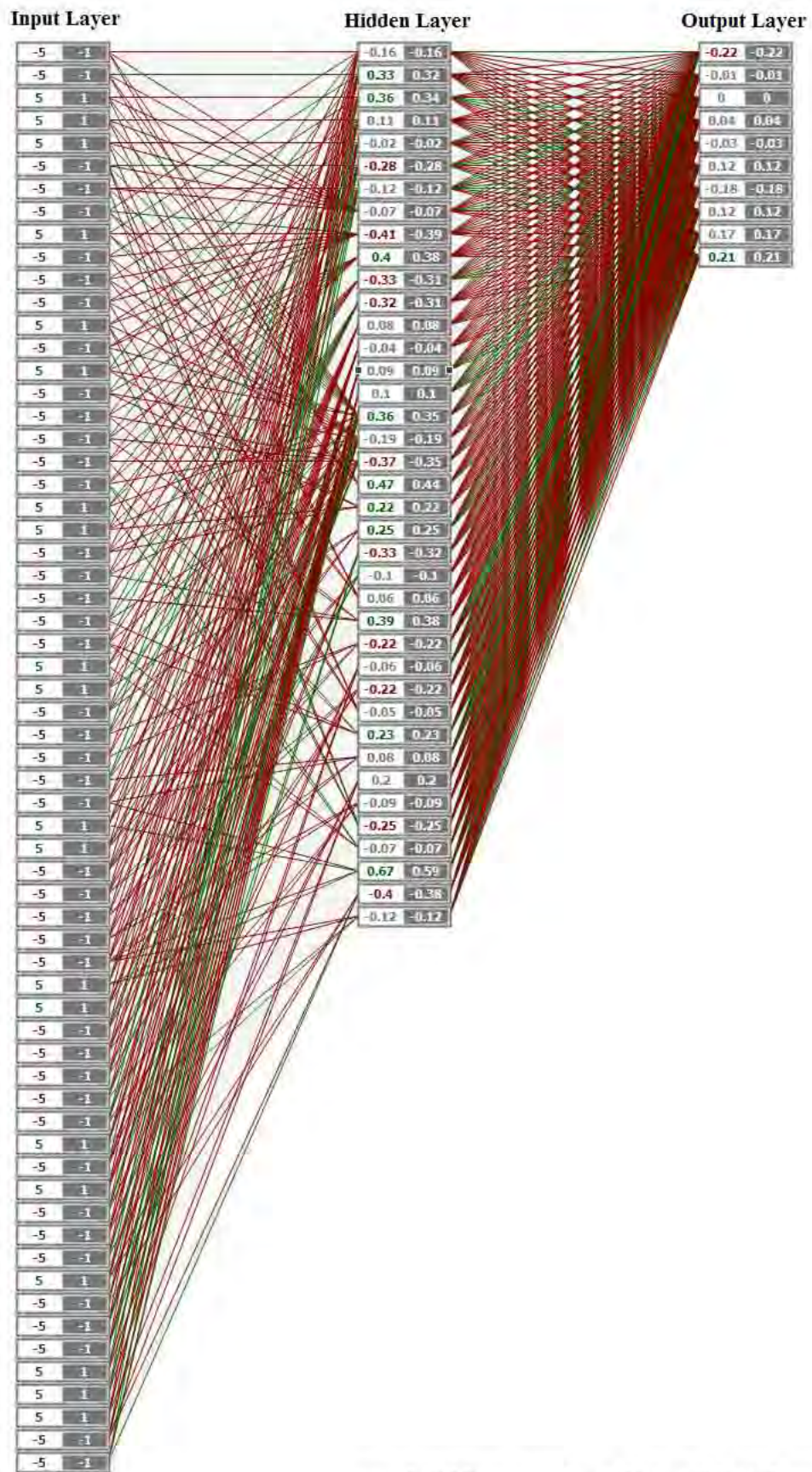


FIGURE C.1: Back Propagation Network with Relative Weights between layers for Ziffer 1

Appendix D

Neural Network Based AR System Work Flow Diagram

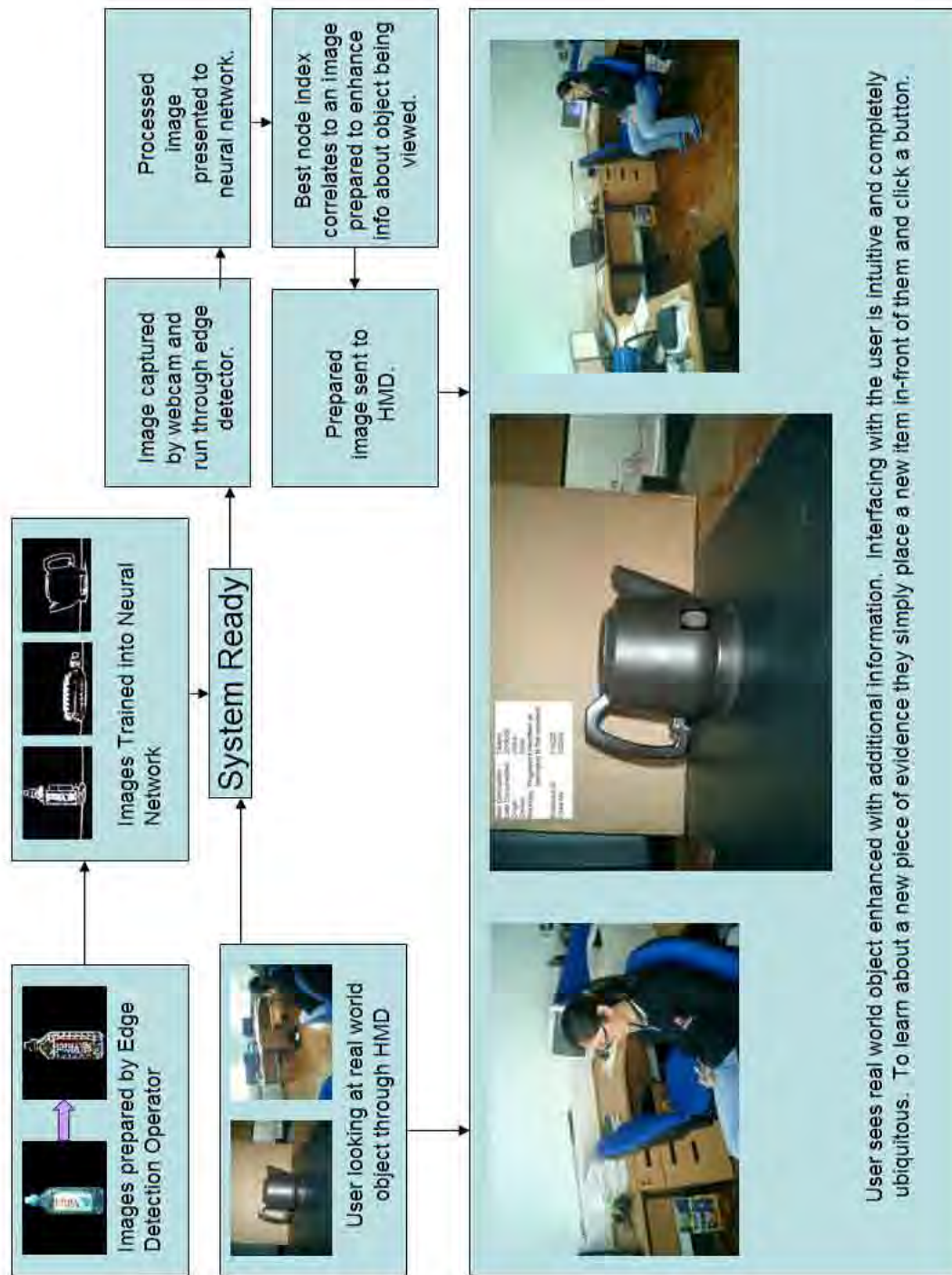


FIGURE D.1

Appendix E

User Consent Form

University of Birmingham

Phd Research Trial

Augmented Reality Experiment to determine the effect that a real world frame of reference has on human memory and perception.



Consent Form

	Please tick to confirm
I have read and understood the instruction sheet.	
I have been given the opportunity to ask questions about the study.	
I am satisfied with the answers given to my queries for this study.	
I understand that I have the right to withdraw from the experiment at any time without explanation.	
I agree to report any eye discomfort or motion sickness that might result from using the Head Mounted Display.	
I agree to the data being collected from me to be used for research purposes.	
I agree to take part in the aforementioned research study.	

Researcher Name Date.....

Participant Name Date.....

FIGURE E.1: Participants each signed a consent form before taking part in trials

Appendix F

Recall Sheet - Experiment 1A

Name..... Experiment No.....

The boxes below are representative of the positions of the objects you have just observed. Please fill in the boxes with the names of the objects you saw in the corresponding position.

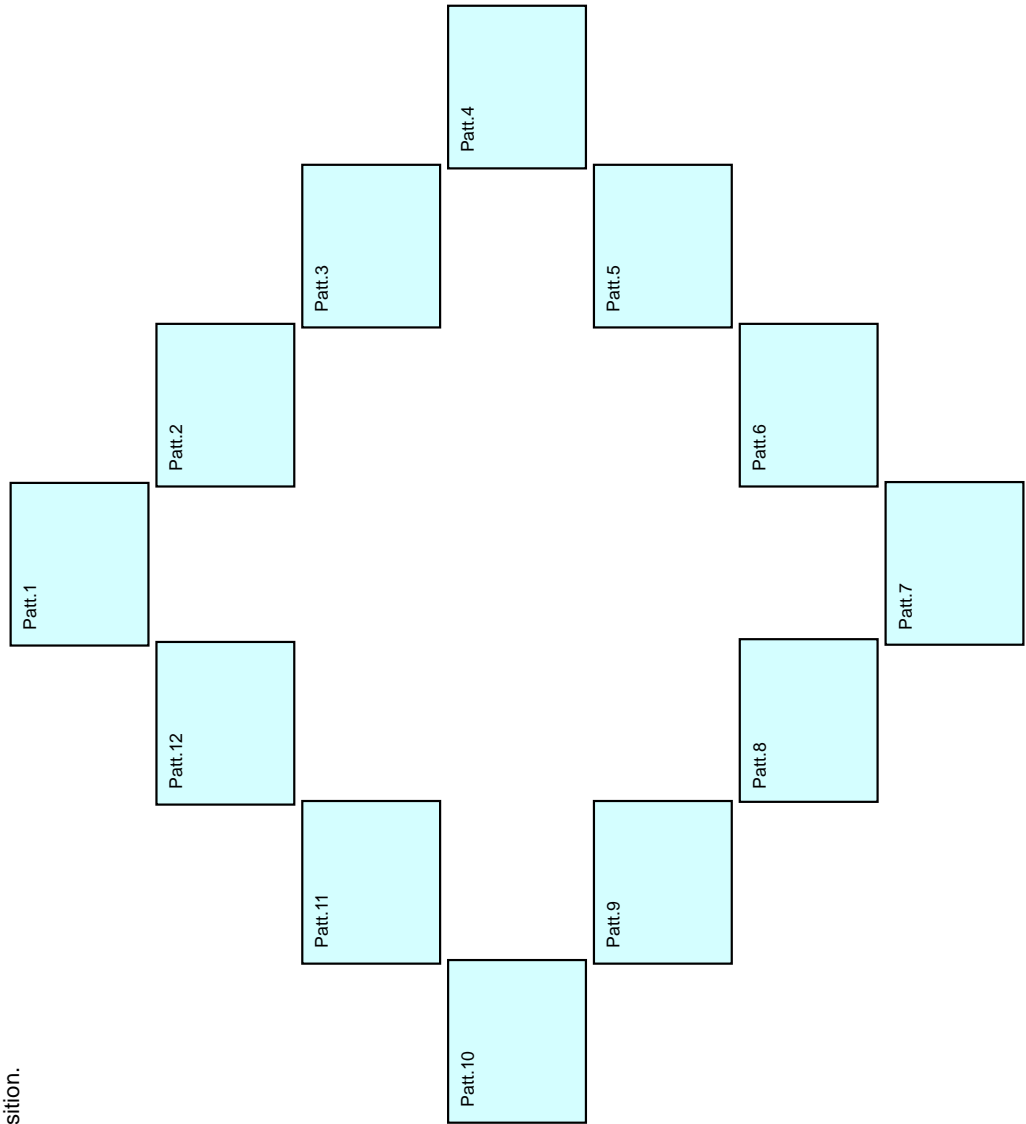


FIGURE F.1: Recall Sheet for Experiment 1A, Real World Frame of Reference

Appendix G

Recall Sheet - Experiment 1B, Blocking Effect

Name..... Experiment No.....

The boxes below are representative of the positions of the objects you have just observed. Please fill in the boxes with the names of the objects you saw in the corresponding position.

Set 1

Set 2

Set 3

Set 4

FIGURE G.1: Recall Sheet for Experiment 1B, Real World Frame of Reference with Blocking Introduced

Appendix H

Recall Sheet - Experiment 2

Name..... Experiment No.....

The boxes below are representative of the positions of the objects you have just observed. Please fill in the boxes with the names of the objects you saw in the corresponding position.

The form consists of 12 light blue rectangular boxes arranged in a 3x4 grid. Each box is labeled with a pattern name, oriented vertically. The labels are as follows:

- Row 1 (top): Patt.6, Patt.2, Patt.1, Patt.8
- Row 2 (middle): Patt.7, Patt.3, Patt.5, Patt.9
- Row 3 (bottom): Patt.10, Patt.11, Patt.12

FIGURE H.1: Recall Sheet for Experiment 2, Immersive Reality Vs Real Life

Appendix I

Experiment 2 - Scan Path Analysis Data

	1	2	3	4	5	6	7	8	9	10	11	12	<i>Node</i>	<i>Status</i>
1	0	1	3	3	8	0	0	0	0	0	7	0	1	4.18
2	4	0	3	0	0	2	6	0	0	0	0	0	3	3.55
3	5	5	0	0	0	0	4	0	0	4	4	0	5	3.55
4	3	0	0	0	6	0	0	3	0	0	0	0	11	3.45
5	5	0	0	3	0	0	0	0	3	0	2	7	2	2.09
6	0	2	0	0	0	0	0	0	0	0	0	0	4	1.82
7	0	0	3	0	0	0	0	0	0	1	0	0	10	1.82
8	0	0	0	2	0	0	0	0	1	0	0	0	12	1.73
9	0	0	0	0	3	0	0	0	0	0	0	1	7	1.36
10	0	0	6	0	0	0	1	0	0	0	2	0	9	0.82
11	7	0	2	0	0	0	0	0	0	6	0	2	8	0.55
12	0	0	0	0	2	0	0	0	1	0	6	0	6	0.36

FIGURE I.1: Augmented reality scan path analysis data, table on right shows average number of visitations per node per test run

	1	2	3	4	5	6	7	8	9	10	11	12	<i>Node</i>	<i>Status</i>
1	0	1	2	2	9	0	0	0	0	0	6	0	1	4.18
2	4	0	4	0	0	3	5	0	0	0	0	0	11	4.0
3	6	4	0	0	0	0	3	0	0	5	3	0	5	3.82
4	4	0	0	0	6	0	0	3	1	0	0	0	3	3.64
5	4	0	0	3	0	0	0	0	2	0	3	8	10	2.27
6	0	3	0	0	0	0	0	0	0	0	0	0	2	2.18
7	0	0	4	0	0	0	0	0	0	3	0	0	12	2.09
8	0	0	0	3	0	0	0	0	2	0	0	0	4	2.0
9	0	0	0	0	3	0	0	0	0	0	0	0	7	1.36
10	0	0	6	0	0	0	0	0	0	0	2	0	9	0.82
11	8	0	3	0	0	0	0	0	0	9	0	1	8	0.73
12	0	0	0	0	4	0	0	0	1	0	9	0	6	0.55

FIGURE I.2: Virtual reality scan path analysis data, table on right shows average number of visitations per node per test run

	1	2	3	4	5	6	7	8	9	10	11	12	<i>Node</i>	<i>Status</i>
1	0	1	3	1	8	0	0	0	0	0	7	0	<i>1</i>	3.91
2	3	0	1	0	0	1	0	0	0	0	0	0	<i>11</i>	3.36
3	4	5	0	0	0	0	2	0	0	1	2	0	<i>5</i>	3.0
4	3	0	0	0	4	0	0	2	1	0	0	0	<i>3</i>	2.55
5	4	0	0	4	0	0	0	0	1	0	2	4	<i>4</i>	1.55
6	0	2	0	0	0	0	0	0	0	0	0	0	<i>12</i>	1.55
7	1	0	2	0	0	0	0	0	0	1	0	0	<i>10</i>	1.36
8	0	0	0	2	0	0	0	0	1	0	0	0	<i>2</i>	1.18
9	0	0	0	0	2	0	0	0	0	0	0	1	<i>7</i>	0.64
10	1	0	4	0	0	0	1	0	0	0	2	0	<i>9</i>	0.55
11	7	0	4	0	0	0	0	0	0	5	0	1	<i>8</i>	0.45
12	0	0	0	0	4	0	0	0	0	0	7	0	<i>6</i>	0.27

FIGURE I.3: Real life scan path analysis data, table on right shows average number of visitations per node per test run

Appendix J

Recall Sheet - Experiment 3

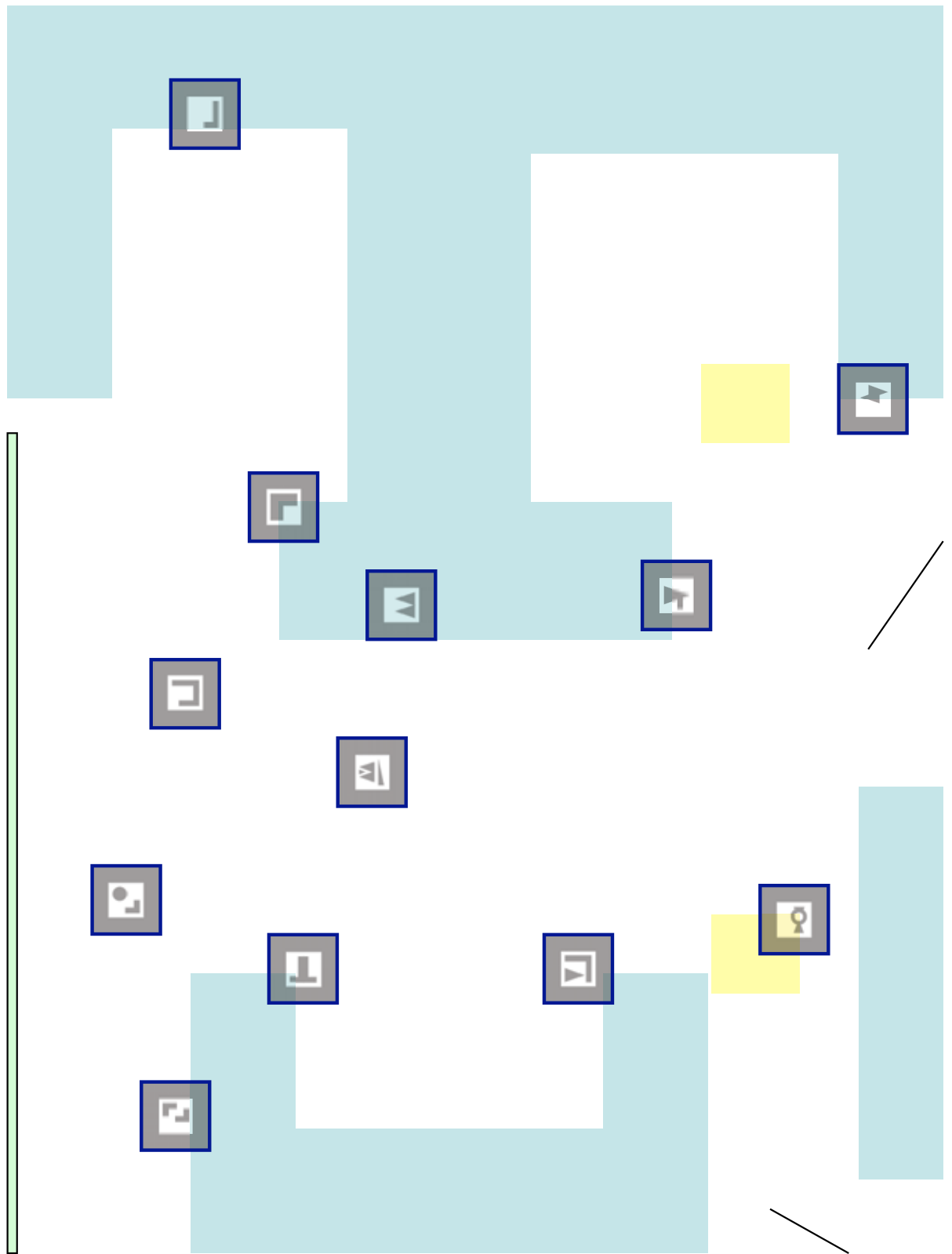


FIGURE J.1: Recall Sheet for Search of a Local Environment laid out as per object in situ on a floor plan of the search area

Bibliography

- [1] M.L. Dertouzos. *What will be: How the new world of information will change our lives*. HarperCollins Publishers, 1997.
- [2] M.A. Livingston, L.J. Rosenblum, S.J. Julier, D. Brown, Y. Baillet, J.E.S. II, J.L. Gabbard, and D. Hix. An augmented reality system for military operations in urban terrain. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2002. NTSA, 2002.
- [3] C. Baber, P. Smith, S. Panesar, F. Yang, and J. Cross. Supporting crime scene investigation. *People and Computers XX, London: Springer-Verlag*, 3:103–116, 2007.
- [4] B.A.J. Fisher. *Techniques of crime scene investigation*. CRC Press, 2004.
- [5] A. Jamieson. A rational approach to the principles and practice of crime scene investigation: I. principles. *Science & Justice*, 44(1):3–7, 2004.
- [6] C. Baber, P. Smith, J. Cross, J. Hunter, and R. McMaster. Crime scene investigation as distributed cognition. *Pragmatics and cognition*, 14:357–385, 2006.
- [7] D. Schofield. Animating and interacting with graphical evidence : Bringing courtrooms to life with virtual reconstructions. *Computer Graphics, Imaging and Visualization, International Conference on*, 0:321–328, 2007. doi: <http://doi.ieeecomputersociety.org/10.1109/CGIV.2007.18>.
- [8] C. McCartney. *Forensic Identification And Criminal Justice: Forensic Science, Justice and Risk*. Willan Publishing, 2006.
- [9] M. Findlay and J. Grix. Challenging Forensic Evidence-Observations on the Use of DNA in Certain Criminal Trials. *Current Issues in Criminal Justice*, 14:269, 2002.

- [10] J.M. Schraagen and H. Leijenhorst. Searching for evidence: Knowledge and search strategies used by forensic scientists. *Linking Expertise and Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum, pages 263–274, 2001.
- [11] M. Weiser. The computer for the 21 st century. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3(3):3–11, 1999.
- [12] K. Kawabata, N. Nishioka, P.C. Lin, H. Nakamura, and H. Kobayashi. Distance measurement method under multiple ultra sonic sensors environment. In *Industrial Electronics, Control, and Instrumentation, 1996., Proceedings of the 1996 IEEE IECON 22nd International Conference on*, volume 2, 1996.
- [13] N.E. Seymour, A.G. Gallagher, S.A. Roman, M.K. O’Brien, V.K. Bansal, D.K. Andersen, and R.M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of Surgery*, 236(4):458, 2002.
- [14] J.C. de Winter, S. de Groot, M. Mulder, P.A. Wieringa, J. Dankelman, and J.A. Mulder. Relationships between driving simulator performance and driving test results. *Ergonomics*, 52(2):137–153, 2009.
- [15] D.L. Roenker, G.M. Cissell, K.K. Ball, V.G. Wadley, and J.D. Edwards. Speed-of-processing and driving simulator training result in improved driving performance. *Human Factors*, 45(2):218, 2003.
- [16] S. Gibson, R.J. Hubbold, J. Cook, and T.L.J. Howard. Interactive reconstruction of virtual environments from video sequences. *Computers & Graphics*, 27(2):293–301, 2003.
- [17] S. Gibson and T.L.J. Howard. Interactive reconstruction of virtual environments from photographs, with application to scene-of-crime analysis. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 41–48. ACM New York, NY, USA, 2000.
- [18] A. Murta, S. Gibson, TLJ Howard, RJ Hubbold, and AJ West. Modelling and rendering for scene of crime reconstruction: A case study. In *Proceedings Eurographics UK*, pages 169–173. Citeseer, 1998.

- [19] T.L.J. Howard, A.D. Murta, and S. Gibson. Virtual environments for scene of crime reconstruction and analysis. In *Proceedings of SPIE*, volume 3960, page 41, 2000.
- [20] R.P. Darken and W.P. Banker. Navigating in natural environments: A virtual environment training transfer study. In *IEEE 1998 Virtual Reality Annual International Symposium, 1998. Proceedings.*, pages 12–19, 1998.
- [21] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Proceedings of Telem manipulator and Telepresence Technologies*, pages 282–292, 1994.
- [22] S. Bryson, C. Levit, N.A.R. Center, and M. Field. The virtual wind tunnel. *IEEE Computer Graphics and Applications*, 12(4):25–34, 1992.
- [23] R. L. Gregory. *Eye and brain: The psychology of seeing*. Toronto: McGraw-Hill, 2 edition, 1973.
- [24] S. Bryson. Virtual reality in scientific visualization. *Commun. ACM*, 39(5):62–71, 1996. ISSN 0001-0782. doi: <http://doi.acm.org/10.1145/229459.229467>.
- [25] A. Jelfs and D. Whitelock. The notion of presence in virtual learning environments: what makes the environment "real" the notion of presence in virtual learning environments: what makes the environment "real" the notion of presence in virtual learning environments: what makes the environment "real". *British Journal of Educational Technology*, 31(2):145–152, 2000.
- [26] R. Stone, D. White, R. Guest, and B. Francis. The virtual scylla: an exploration of 'serious games', artificial life and simulation complexity. *Virtual Reality*, 13(1): 13–25, 2009.
- [27] W. Piekarski, B. Avery, B.H. Thomas, and P. Malbezin. Integrated head and hand tracking for indoor and outdoor augmented reality. In *IEEE Virtual Reality Conference*, pages 11–276, 2004.
- [28] P. Felkel, A. Fuhrmann, A. Kanitsar, and R. Wegenkittl. Surface reconstruction of the branching vessels for augmented reality aided surgery. accepted to biosignal 2002. In *The 16th international EURASIP conference. It will be held in Brno, Czech Republic, June, 2002*.

- [29] J. Traub, M. Feuerstein, M. Bauer, EU Schirmbeck, H. Najafi, R. Bauernschmitt, and G. Klinker. Augmented reality for port placement and navigation in robotically assisted minimally invasive cardiovascular surgery. In *International Congress Series*, volume 1268, pages 735–740. Elsevier, 2004.
- [30] R. Azuma. A survey of augmented reality. *Presence*, 6:355–385, 1995.
- [31] W. Piekarski and B. Thomas. Arquake: the outdoor augmented reality gaming system. 2002.
- [32] B. Thomas, N. Krul, B. Close, and W. Piekarski. Usability and playability issues for arquake. In *Int'l Workshop on Entertainment Computing*, pages 455–462, 2002.
- [33] M. Wellner, A. Schaufelberger, J. Zitzewitz, and R. Riener. Evaluation of visual and auditory feedback in virtual obstacle walking. *Presence: Teleoperators and Virtual Environments*, 17(5):512–524, 2008.
- [34] E. Bennett and B. Stevens. The effect that the visual and haptic problems associated with touching a projection augmented model have on object-presence. *Presence: Teleoper. Virtual Environ.*, 15(4):419–437, 2006. ISSN 1054-7460. doi: <http://dx.doi.org/10.1162/pres.15.4.419>.
- [35] M.O. Ernst and H.H. Bühlhoff. Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–169, 2004.
- [36] R.A. Jacobs. What determines visual cue reliability? *Trends in Cognitive Sciences*, 6(8):345–350, 2002. doi: 10.1016/S1364-6613(02)01948-4. URL [http://dx.doi.org/10.1016/S1364-6613\(02\)01948-4](http://dx.doi.org/10.1016/S1364-6613(02)01948-4).
- [37] J.E. Cutting. How the eye measures reality and virtual reality. *Behavior Research Methods Instruments and Computers*, 29:27–36, 1997.
- [38] X. Wang, R. Chen, and R. Wang. A cognitive study on the effectiveness of an augmented virtuality-based collaborative design space. In *CDVE '08: Proceedings of the 5th international conference on Cooperative Design, Visualization, and Engineering*, pages 253–256, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-88010-3. doi: http://dx.doi.org/10.1007/978-3-540-88011-0_36.
- [39] A.L. Brooks and E. Petersson. Play Therapy Utilizing the Sony EyeToy. *Presence: Teleoperators and Virtual Environments*, 2005.

- [40] H. T. Regenbrecht, M. Wagner, and G. Baratoff. Magicmeeting: A collaborative tangible augmented reality system. *Virtual Reality*, 6(3):151–166, 2002. URL <http://dx.doi.org/10.1007/s100550200016>.
- [41] H. Regenbrecht, C. Ott, M. Wagner, T. Lum, P. Kohler, W. Wilke, and E. Mueller. An augmented virtuality approach to 3d videoconferencing. *Mixed and Augmented Reality, IEEE / ACM International Symposium on*, 0:290, 2003. doi: <http://doi.ieeecomputersociety.org/10.1109/ISMAR.2003.1240725>.
- [42] D. Wagner and D. Schmalstieg. Artoolkitplus for pose tracking on mobile devices. In *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)*, pages 139–146, 2007.
- [43] G. Reitmayr and T. Drummond. Going out: robust model-based tracking for outdoor augmented reality. *Mixed and Augmented Reality, IEEE / ACM International Symposium on*, 0:109–118, 2006. doi: <http://doi.ieeecomputersociety.org/10.1109/ISMAR.2006.297801>.
- [44] D. Beier, R. Billert, B. Bruderlin, D. Stichling, and B. Kleinjohann. Marker-less vision based tracking for mobile augmented reality. In *ISMAR '03: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 258, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-2006-5.
- [45] W. Pasman, C. Woodward, M. Hakkarainen, P. Honkamaa, and J. Hyvakka. Augmented reality with large 3d models on a pda: implementation, performance and use experiences. In *VRCAI '04: Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry*, pages 344–351, New York, NY, USA, 2004. ACM. ISBN 1-58113-884-9. doi: <http://doi.acm.org/10.1145/1044588.1044663>.
- [46] D.J. Weintraub and M. Ensing. *Human Factors Issues in Head-Up Display Design: The Book of HUD*. Crew System Ergonomics, 1992.
- [47] C.D. Wickens, J.D. Lee, Y. Liu, and Becker S.E. *An Introduction to Human Factors Engineering*. Prentice Hall, 2 edition, 1997.
- [48] T.T. Elvins. Augmented reality: “the future’s so bright, i gotta wear (see-through) shades”. *ACM SIGGRAPH Computer Graphics*, 32(1):11–13, 1998.

- [49] R.T. Azuma. Augmented reality: Approaches and technical challenges. *Fundamentals of wearable computers and augmented reality*, page 27, 2001.
- [50] S. Hecht and E.U. Mintz. The visibility of single lines at various illuminations and the retinal basis of visual resolution. *J. Gen. Physiol*, 22:593–612, 1939.
- [51] J.P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays. In *in Medical Visualization. Presence: Teleoperators and Virtual Environments*, pages 287–309, 2000.
- [52] J.P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence: Teleoperators and Virtual Environments*, 11(6):610–625, 2002.
- [53] T. Hollerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway. Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics*, 23(6):779–785, 1999.
- [54] G. Beach, C.J. Cohen, J. Braun, and G. Moody. Eye tracker system for use with head mounted displays. In *1998 IEEE International Conference on Systems, Man, and Cybernetics, 1998*, volume 5, 1998.
- [55] M.C. Amann, T. Bosch, M. Lescure, R. Myllylä, and M. Rioux. Laser ranging: a critical review of usual techniques for distance measurement. *Optical Engineering*, 40:10, 2001.
- [56] W.E.L. Grimson, G.J. Ettinger, S.J. White, P.L. Gleason, T. Lozano-Pérez, W.M. Wells III, and R. Kikinis. Evaluating and validating an automated registration system for enhanced reality visualization in surgery. In *Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 3–12. Springer-Verlag London, UK, 1995.
- [57] G. Gordon, M. Billinghurst, M. Bell, J. Woodfill, B. Kowalik, A. Erendi, J. Tilander, T. Inc, and C.A. Palo Alto. The use of dense stereo range data in augmented reality. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, pages 14–23, 2002.

- [58] G. H. Alusi, A. C. Tan, A. D. Linney, K. Raof, and A. Wright. Three dimensional tracking with ultrasound for augmented reality applications in skull base surgery. In *CVRMed-MRCAS '97: Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*, pages 511–517, London, UK, 1997. Springer-Verlag. ISBN 3-540-62734-0.
- [59] A. State, G. Hirota, D.T. Chen, W.F. Garrett, and M.A. Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 429–438, New York, NY, USA, 1996. ACM. ISBN 0-89791-746-4. doi: <http://doi.acm.org/10.1145/237170.237282>.
- [60] M.A. Livingston. Magnetic tracker calibration for improved augmented reality registration. *Teleoperators and Virtual Environments*, 1997.
- [61] A.M. Shkel, R. Horowitz, A.A. Seshia, S. Park, and R.T. Howe. Dynamics and control of micromachined gyroscopes. In *American Control Conference, 1999. Proceedings of the 1999*, volume 3, 1999.
- [62] J. Geen and D. Krakauer. New imems® angular-rate-sensing gyroscope. *Analog Dialogue*, 37(3):1–4, 2003.
- [63] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan. Real-time vision-based camera tracking for augmented reality applications. In *ACM Symposium on Virtual Reality Software and Technology*, pages 87–94. ACM Press, 1997.
- [64] Y. Cho and U. Neumann. Multiring fiducial systems for scalable fiducial-tracking augmented reality. *Presence: Teleoperators and Virtual Environments*, 10(6):599–612, 2001. doi: 10.1162/105474601753272853. URL <http://www.mitpressjournals.org/doi/abs/10.1162/105474601753272853>.
- [65] T. Ohshima, K Satoh, H. Yamamoto, and H. Tamura. Arhockey: A case study of collaborative augmented reality. In *Proc. IEEE VRAIS '98*, pages 268–275, 1998.
- [66] V. Ferrari, T. Tuytelaars, and L. Van Gool. Markerless augmented reality with a real-time affine region tracker. In *IEEE and ACM International Symposium on Augmented Reality, 2001. Proceedings*, pages 87–96, 2001.

- [67] J.E. Swan, A. Jones, E. Kolstad, M.A. Livingston, and H.S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):429–442, 2007. ISSN 1077-2626. doi: <http://doi.ieeecomputersociety.org/10.1109/TVCG.2007.1035>.
- [68] D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc. New York, NY, USA, 1982.
- [69] J.R. Wilson and M. D’Cruz. Virtual and interactive environments for work of the future. *International Journal of Human-Computer Studies*, 64(3):158–169, 2006.
- [70] B.E. Riecke. Consistent left-right reversals for visual path integration in virtual reality: More than a failure to update one’s heading? *Presence: Teleoperators and Virtual Environments*, 17(2):143–175, 2008.
- [71] R.A. Ruddle, S.J. Payne, and D.M. Jones. Navigating large-scale “desk-top” virtual buildings: Effects of orientation aids and familiarity. *Presence: Teleoper. Virtual Environ.*, 7(2):179–192, 1998. ISSN 1054-7460. doi: <http://dx.doi.org/10.1162/105474698565668>.
- [72] D.A. Bowman and L.F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM New York, NY, USA, 1997.
- [73] R.A. Ruddle and D.M. Jones. Movement in cluttered virtual environments. *Presence: Teleoper. Virtual Environ.*, 10(5):511–524, 2001. ISSN 1054-7460. doi: <http://dx.doi.org/10.1162/105474601753132687>.
- [74] S.S. Chance, F. Gaunet, P.M. Berthelot, A.C. Beall, and J.M. Loomis. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *presence: Teleoperators and virtual environments*. *Presence*, 7:168–178, 1998.
- [75] R.A. Ruddle and S. Lessels. For efficient navigational search, humans require full physical movement, but not a rich visual scene. *Psychological Science*, 17(6):460–465, 2006. ISSN 0956-7976. doi: 10.1111/j.1467-9280.2006.01728.x. URL <http://dx.doi.org/10.1111/j.1467-9280.2006.01728.x>.

- [76] C.G.L. Cao and P. Milgram. Direction and location are not sufficient for navigating in nonrigid environments: An empirical study in augmented reality. *Presence: Teleoperators and Virtual Environments*, 16(6):584–602, 2007.
- [77] 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*, 17(1):91–101, 2008.
- [78] J.P. Rolland, C.A. Burbeck, W. Gibson, and D. Ariely. Towards quantifying depth and size perception in 3D virtual environments. *Presence: Teleoperators and Virtual Environments*, 4(1):24–48, 1995.
- [79] S.R. Ellis and B.M. Menges. Localization of virtual objects in the near visual field. *Human Factors*, 40(3):415–416, 1998.
- [80] R.L. Woods, I. Fetchenheuer, F. Vargas-martn, and E. Peli. The impact of non-immersive head-mounted displays (hmd) on the visual field. *J Soc Inform Display*, 2003.
- [81] F. Tong, K. Nakayama, J.T. Vaughan, and N. Kanwisher. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21:753–759, 1998.
- [82] T. Yamazoe, S. Kishi, T. Shibata, T. Kawai, and M. Otsuki. Reducing binocular rivalry in the use of monocular head-mounted display. *Display Technology, Journal of*, 3, 2007.
- [83] R.S. Laramée and C. Ware. Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9:238–251, 2002.
- [84] K.R. Boff and J.E. Lincoln. Engineering Data Compendium: Human Perception and Performance. AAMRL, WPAFB, Ohio, Volumes 1-3, 1988.
- [85] M.A. Livingston, C. Zanbaka, J.E.S. II, and H.S. Smallman. Objective measures for the effectiveness of augmented reality. *Proceedings of IEEE Virtual Reality (Poster Session)*, 2005.
- [86] M.A. Livingston. Quantification of visual capabilities using augmented reality displays. In *IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006*, pages 3–12, 2006.

- [87] F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3):551–566, 1968.
- [88] B. Shneiderman. Direct manipulation: A step beyond programming. *IEEE Computer*, 16(8):57–69, 1983.
- [89] J. Zhang and D.A. Norman. Representations in distributed cognitive tasks. *Cognitive Science*, 18:87–122, 1994.
- [90] D. Kirsh. The intelligent use of space. *Artificial Intelligence*, 73:31–68, 1995.
- [91] F.P. Brooks. What’s real about virtual reality? *IEEE Comput. Graph. Appl.*, 19(6):16–27, 1999. ISSN 0272-1716. doi: <http://dx.doi.org/10.1109/38.799723>.
- [92] M. Tavanti and M. Lind. 2D vs 3D, implications on spatial memory. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001*, pages 139–145, 2001.
- [93] A. Cockburn. Revisiting 2d vs 3d implications on spatial memory. In *Proceedings of the fifth conference on Australasian user interface-Volume 28*, pages 25–31. Australian Computer Society, Inc. Darlinghurst, Australia, Australia, 2004.
- [94] M. St John, M.B. Cowen, H.S. Smallman, and H.M. Oonk. The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, 43(1):79, 2001.
- [95] J. Lehtikoinen and R. Suomela. Accessing context in wearable computers. *Personal and Ubiquitous Computing*, 6(1):64–74, 2002.
- [96] R. Banks and C. D. Wickens. Commanders display of terrain information: Manipulations of display dimensionality and frame of reference to support battlefield visualization. In *Army Research Laboratory Final Report*, 1999.
- [97] K. Mania, T. Troscianko, R. Hawkes, and A. Chalmers. Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence: Teleoperators & Virtual Environments*, 12(3):296–310, 2003.
- [98] K. Mania, A. Robinson, and K.R. Brandt. The effect of memory schemas on object recognition in virtual environments. *Presence: Teleoperators & Virtual Environments*, 14(5):606–615, 2005.

- [99] H.Q. Dinh, N. Walker, C. Song, A. Kobayashi, and L.F. Hodges. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings of the IEEE Virtual Reality*, volume 99, pages 222–228. Citeseer, 1999.
- [100] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188–194, 2005.
- [101] M. Land, N. Mennie, and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. *PERCEPTION-LONDON-*, 28(11):1311–1328, 1999.
- [102] N. Moray, M. Fitter, D. Ostry, D. Favreau, and V. Nagy. Attention to pure tones. *The Quarterly Journal of Experimental Psychology*, 28(2):271–283, 1976.
- [103] W.H. Teichner and M.J. Krebs. Laws of visual choice reaction time. *Psychol Rev*, 81(1):75–98, 1974.
- [104] W.F. Bacon and H.E. Egeth. Local processes in preattentive feature detection. *J Exp Psychol Hum Percept Perform*, 17(1):77–90, 1991.
- [105] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.
- [106] A. Treisman and S. Sato. Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3):459–478, 1990.
- [107] R.E. Christ. Review and analysis of color coding research for visual displays(in aircraft). *Human Factors*, 17:542–570, 1975.
- [108] M.R. Beck, M.S. Peterson, and M. Vomela. Memory for where, but not what, is used during visual search. *Journal of Experimental Psychology Human Perception and Performance*, 32(2):235, 2006.
- [109] F.B.R. Parmentier, G. Elford, and M. Maybery. Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology Learning Memory and Cognition*, 31(3):412, 2005.
- [110] G.A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review*, 63(2):81–97, 1956.

-
- [111] C.D. Wickens and J.G. Hollands. *Engineering Psychology and Human Performance*. ISBN: 0-321-04711-7, 2000.
- [112] U. Neumann, S. You, Y. Cho, J. Lee, and J. Park. Augmented reality tracking in natural environments. In *International Symposium on Mixed Realities*, volume 24, 1999.
- [113] L.A. MacVittie. *XAML in a Nutshell*. O'Reilly Media, Inc., 2006.
- [114] C. Sells and I. Griffiths. *Programming Windows presentation foundation*. O'Reilly Media, Inc., 2005.
- [115] K. Rehman. Augmented reality in support of interaction for location-aware applications. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 286–287. IEEE Computer Society Washington, DC, USA, 2004.
- [116] H.W.P. Beadle, B. Harper, G.Q. Maguire, and J. Judge. Location aware mobile computing. In *Proc. ICT'97 (IEEE/IEE Int. Conf. on Telecomm.)*, 1997.
- [117] M. Wagner. Building wide-area applications with the ar toolkit. In *The First IEEE International Augmented Reality Toolkit Workshop*, 2002.
- [118] B.J. Van der Zwaag, C.H. Slump, and L. Spaanenburg. Extracting knowledge from supervised neural networks in image processing. 2003.
- [119] B.J. van der Zwaag, C. Slump, and L. Spaanenburg. Analysis of neural networks for edge detection. In *Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing*, pages 28–29, 2002.
- [120] J.J. Clark. Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):43–57, 1989.
- [121] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, pages 679–698, 1986.
- [122] M. Sharifi, M. Fathy, and M.T. Mahmoudi. A classified and comparative study of edge detection algorithms. In *Information Technology: Coding and Computing, 2002. Proceedings. International Conference on*, pages 117–120, 2002.

- [123] S. Watanabe. *Pattern recognition: human and mechanical*. John Wiley & Sons, Inc. New York, NY, USA, 1985.
- [124] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):4–37, 2000.
- [125] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE transactions on pattern analysis and machine intelligence*, 15(10):1042–1052, 1993.
- [126] K.S. Fu. Syntactic pattern recognition. *Applications of Pattern Recognition*, page 37, 1982.
- [127] C.M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, USA, 1995.
- [128] A. Martínez-Estudillo, F. Martínez-Estudillo, C. Hervás-Martínez, and N. García-Pedrajas. Evolutionary product unit based neural networks for regression. *Neural Networks*, 19(4):477–486, 2006.
- [129] L.O. Chua and L. Yang. Cellular neural networks: Applications. *IEEE Transactions on Circuits and Systems*, 35(10):1273–1290, 1988.
- [130] E.G. Rajan, V. Kumar, A.G.S. Kiran, A. Chowdhary, and N.S. Murthy. Neural automata based object recognition. In *IEEE International Conference on Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century.*, volume 2, 1995.
- [131] E.D. Sontag and H.J. Sussmann. Backpropagation can give rise to spurious local minima even for networks without hidden layers. *Complex Systems*, 3(1):91–106, 1989.
- [132] A.J. Annema. *Feed-Forward Neural Networks: Vector Decomposition Analysis, Modelling, and Analog Implementation*. Springer, 1995.
- [133] B.A. Pearlmutter and R. Rosenfeld. Chaitin-kolmogorov complexity and generalization in neural networks. In *Proceedings of the 1990 conference on Advances in neural information processing systems 3 table of contents*, pages 925–931. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1990.

- [134] M. Fiala. Comparing artag and artoolkit plus fiducial marker systems. *Haptic Audio Visual Environments and their Applications, 2005. IEEE International Workshop*, 2005.
- [135] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L.M. Encarnação, M. Gervautz, and W. Purgathofer. The studierstube augmented reality project. *Presence: Teleoperators and Virtual Environments*, 11(1):33–54, 2002.
- [136] M. Fiala. Artag, an improved marker system based on artoolkit. *NRC, CNRC Institute for Information Technology*, 2004.
- [137] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 2, 2005.
- [138] W. Mou and T.P. McNamara. Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology Human Perception and Performance*, 28(1):162–170, 2002.
- [139] E.M. Altmann. Memory in chains: Modeling primacy and recency effects in memory for order. In *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society: August 13-15, 2000, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA*, page 31. Lawrence Erlbaum Associates, 2000.
- [140] J.S. Miller, D.L. McKinzie, K.S. Kraebel, and N.E. Spear. Changes in the expression of stimulus selection: Blocking represents selective memory retrieval rather than selective associations. *Learning and Motivation*, 27(3):307–316, 1996.
- [141] K. Lemos, T. Schnell, T. Etherington, T. Vogl, and A. Postikov. Synthetic vision systems: human performance assessment of the influence of terrain density and texture. In *Digital Avionics Systems Conference, 2003. DASC'03. The 22nd*, volume 2, 2003.
- [142] J.F. Knight and C. Baber. Wearable computers and the possible development of musculoskeletal disorders. In *Wearable Computers, 2000. The Fourth International Symposium on*, pages 171–172, 2000.

-
- [143] M.I. Benta. Studying communication networks with Agna 2.1. *Cognition, Brain, Behavior*, 9(3):567–574, 2005.
- [144] C. Baber, J. Knight, D. Haniff, and L. Cooper. Ergonomics of wearable computers. *Mobile Networks and Applications*, 4(1):15–21, 1999.
- [145] J.F. Knight and C. Baber. Effect of head-mounted displays on posture. *Human factors*, 49(5):797, 2007.
- [146] AC Boud, C. Baber, and SJ Steiner. Virtual reality: A tool for assembly? *Presence: Teleoperators & Virtual Environments*, 9(5):486–496, 2000.
- [147] M. Rucci and A. Casile. Fixational instability and natural image statistics: Implications for early visual representations. *Network: Computation in Neural Systems*, 16(2):121–138, 2005.