

# Loughborough University Institutional Repository

---

## *Spatial integration in computer-augmented realities*

This item was submitted to Loughborough University's Institutional Repository by the/an author.

**Additional Information:**

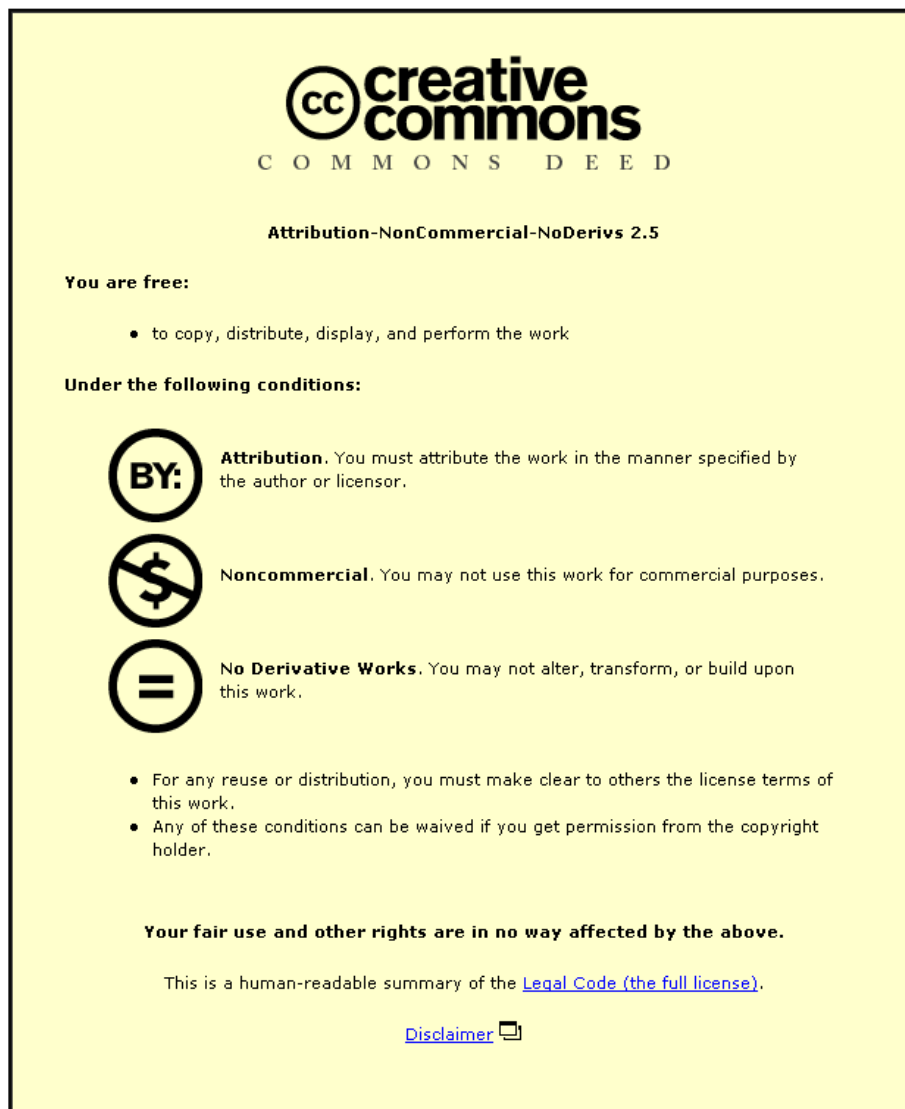
- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

**Metadata Record:** <https://dspace.lboro.ac.uk/2134/14745>

**Publisher:** © Eric William Tatham

Please cite the published version.

This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (<https://dspace.lboro.ac.uk/>) under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>



**University Library**

Author/Filing Title ..... TATHAM .....

Class Mark ..... T .....

Please note that fines are charged on ALL  
overdue items.

|  |  |  |
|--|--|--|
|  |  |  |
|--|--|--|

0402939468







**Spatial Integration in Computer-Augmented Realities**

by

*Eric William Tatham*

Doctoral Thesis


Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

(30 June 2003)

© by Eric William Tatham, 2003

|  |
|--|
|  Loughborough<br>University<br>Physical Library |
| Date Sept 04   |
| Class  |
| Acc No. 040 293946   |

---

|   |           |
|---|-----------|
| <b>List of figures</b>  | .....vi   |
| <b>Acknowledgements</b>                                       | .....xii  |
| <b>Abstract</b>   | .....xiii |
| <br>  |           |
| <b>1.0 Introduction</b>                                       | ..... 1   |
| 1.1 From virtual to augmented reality                         | ..... 1   |
| 1.2 Suspending disbelief                                      | ..... 3   |
| 1.3 Historical background                                     | ..... 3   |
| 1.4 Optical see through versus video see through              | ..... 6   |
| 1.5 Motivation  | ..... 7   |
| 1.6 Implementation hurdles                                    | ..... 7   |
| 1.7 Thesis statement  | ..... 8   |
| <br>  |           |
| <b>2.0 Potential applications of augmented reality</b>        | ..... 12  |
| 2.1 Manufacturing and maintenance                             | ..... 12  |
| 2.2 Construction  | ..... 14  |
| 2.3 Medical imaging   | ..... 16  |
| 2.4 Annotating the real world                                 | ..... 17  |
| 2.5 Education and training                                    | ..... 19  |
| 2.6 Telerobotics  | ..... 19  |
| 2.7 Low vision aids.  | ..... 20  |
| 2.8 Design, visualisation and collaborative working           | ..... 21  |
| 2.9 Concluding remarks  | ..... 23  |
| <br>  |           |
| <b>3.0 Technical issues in implementing augmented reality</b> | ..... 24  |
| 3.1 Registration and tracking                                 | ..... 24  |
| 3.1.1 <i>Mechanical linkage</i>                               | ..... 26  |
| 3.1.2 <i>Magnetic field sensing</i>                           | ..... 27  |
| 3.1.3 <i>Optical sensing</i>                                  | ..... 27  |
| 3.1.4 <i>Acoustic time of flight</i>                          | ..... 27  |
| 3.1.5 <i>Inertial transducers</i>                             | ..... 28  |

---

|            |   |       |           |
|------------|---|-------|-----------|
| 3.1.6      | <i>GPS</i>  | ..... | 28        |
| 3.1.7      | <i>Hybrid approaches</i>                              | ..... | 28        |
| 3.2        | Interactions  | ..... | 29        |
| 3.2.1      | <i>Collisions</i>                                     | ..... | 29        |
| 3.2.2      | <i>Light interactions</i>                             | ..... | 30        |
| 3.2.3      | <i>Real-world modelling requirements</i>              | ..... | 34        |
| 3.3        | Displays  | ..... | 35        |
| 3.3.1      | <i>Video-see-through display</i>                      | ..... | 35        |
| 3.3.2      | <i>Optical-see-through display</i>                    | ..... | 36        |
| 3.3.3      | <i>Monitor-based display</i>                          | ..... | 38        |
| 3.3.4      | <i>Projector-based display</i>                        | ..... | 38        |
| 3.3.5      | <i>Optical versus video see through display</i>       | ..... | 39        |
| 3.4        | Discussion  | ..... | 42        |
| <b>4.0</b> | <b>Perceptual issues in augmented reality</b>         | ..... | <b>43</b> |
| 4.1        | Depth perception overview                             | ..... | 43        |
| 4.1.1      | <i>Convergence</i>                                    | ..... | 44        |
| 4.1.2      | <i>Binocular disparity</i>                            | ..... | 44        |
| 4.1.3      | <i>Motion cues</i>                                    | ..... | 45        |
| 4.1.4      | <i>Accommodation</i>                                  | ..... | 46        |
| 4.1.5      | <i>Interposition</i>                                  | ..... | 46        |
| 4.1.6      | <i>Illumination and colour gradients</i>              | ..... | 47        |
| 4.1.7      | <i>Perspective cues</i>                               | ..... | 48        |
| 4.2        | Depth perception in pictures and virtual environments | ..... | 50        |
| 4.3        | Concluding remarks                                    | ..... | 52        |
| <b>5.0</b> | <b>The estimation of depth in augmented reality</b>   | ..... | <b>54</b> |
| 5.1        | Experiment aim  | ..... | 54        |
| 5.2        | Experiment hypothesis                                 | ..... | 55        |
| 5.3        | Previous virtual depth measurement experiments        | ..... | 55        |
| 5.4        | Comment on earlier experiments                        | ..... | 57        |

---

|            |  |                 |
|------------|--|-----------------|
| 5.5        | Apparatus modifications  | .....57         |
| 5.6        | Test images  | .....61         |
| 5.7        | Disparity adjustment   | .....65         |
| 5.8        | Participants   | .....66         |
| 5.9        | Additional precautions   | .....67         |
|            | 5.9.1 <i>Experimental equipment and environment</i>                | .....67         |
|            | 5.9.2 <i>Visual anomalies</i>                                      | .....67         |
|            | 5.9.3 <i>Practice effects</i>                                      | .....70         |
| 5.10       | Procedure  | .....71         |
| 5.11       | Results  | .....73         |
| 5.12       | Analysis   | .....74         |
| 5.13       | Discussion   | .....80         |
| <b>6.0</b> | <b>Compositing images in optical see through augmented reality</b> | <b>.....82</b>  |
| 6.1        | The alpha channel  | .....82         |
| 6.2        | Methods for applying an alpha channel to real scenes               | .....84         |
| 6.3        | A mathematical model for see through augmented reality             | .....85         |
| 6.4        | Error effects  | .....93         |
| 6.5        | Conclusion   | .....100        |
| <b>7.0</b> | <b>Alpha channel masking in optical see through displays</b>       | <b>.....101</b> |
| 7.1        | Transmission approach  | .....101        |
|            | 7.1.1 <i>Building a prototype</i>                                  | .....103        |
|            | 7.1.2 <i>Compositing demonstration</i>                             | .....106        |
|            | 7.1.3 <i>Demonstration using a real scene</i>                      | .....121        |
|            | 7.1.4 <i>Transmission masking discussion</i>                       | .....127        |
| 7.2        | Projection approach  | .....129        |
|            | 7.2.1 <i>Projection masking demonstration</i>                      | .....130        |
|            | 7.2.2 <i>Projection masking discussion</i>                         | .....133        |
| 7.3        | Reflection approach  | .....134        |
| 7.4        | Discussion   | .....138        |

---

|            |  |                 |
|------------|--|-----------------|
| <b>8.0</b> | <b>Scene reconstruction</b>                                      | <b>.....139</b> |
| 8.1        | Passive sensing  | .....139        |
|            | 8.1.1 <i>Shape from shading</i>                                  | .....140        |
|            | 8.1.2 <i>Shape from silhouettes</i>                              | .....140        |
|            | 8.1.3 <i>Shape from stereo or motion</i>                         | .....141        |
|            | 8.1.4 <i>Shape from photo consistency</i>                        | .....142        |
| 8.2        | Active sensing   | .....143        |
| 8.3        | Texture acquisition  | .....145        |
| 8.4        | Conclusion   | .....145        |
| <br>       |  |                 |
| <b>9.0</b> | <b>Common illumination</b>                                       | <b>.....146</b> |
| 9.1        | Illumination modelling   | .....147        |
| 9.2        | Radiosity rendering  | .....150        |
|            | 9.2.1 <i>Progressive refinement</i>                              | .....156        |
|            | 9.2.2 <i>Substructuring and adaptive subdivision</i>             | .....158        |
|            | 9.2.3 <i>Hierarchical radiosity</i>                              | .....159        |
| 9.3        | Previous common illumination approaches                          | .....161        |
| 9.4        | Finding a plausible common illumination solution                 | .....164        |
|            | 9.4.1 <i>Theoretical basis</i>                                   | .....164        |
|            | 9.4.2 <i>Solving the equations</i>                               | .....170        |
|            | 9.4.3 <i>Solving Quadratic Programming problems</i>              | .....175        |
| 9.5        | Viewing and rendering software implementation                    | .....176        |
| 9.6        | Testing  | .....177        |
|            | 9.6.1 <i>Constructing the enclosure</i>                          | .....179        |
|            | 9.6.2 <i>The illumination inference algorithm</i>                | .....181        |
|            | 9.6.3 <i>A synthesised example</i>                               | .....182        |
|            | 9.6.4 <i>Determining surface radiosities within a real scene</i> | .....189        |
| 9.7        | Discussion   | .....194        |

---

|   |          |
|---|----------|
| <b>10.0 Discussion and conclusion</b>                                   | .....196 |
| 10.1 Discussion   | .....196 |
| 10.1.1 <i>Modelling real-world geometry</i>                             | .....198 |
| 10.1.2 <i>Modelling material properties</i>                             | .....198 |
| 10.1.3 <i>Maintaining registration</i>                                  | .....199 |
| 10.1.4 <i>Displaying real-virtual composites</i>                        | .....199 |
| 10.2 Conclusion   | .....200 |
| <br>  |          |
| <b>Bibliography</b>   | .....207 |
| <br>  |          |
| <b>Appendix A</b> <i>(Camera calibration)</i>                           | .....221 |
| <b>Appendix B</b> <i>(Instructions to experiment participants)</i>      | .....224 |
| <b>Appendix C</b> <i>(Depth experiment result sheets)</i>               | .....226 |
| <b>Appendix D</b> <i>(Provisional patent application)</i>               | .....229 |
| <b>Appendix E</b> <i>(Papers based on work described in the thesis)</i> | .....239 |
| <b>Appendix F</b> <i>(Software CD-ROM)</i>                              | .....267 |

---

**List of figures**

- Figure 1.1 ..... The Mixed Reality Continuum*
- Figure 1.2 ..... Pepper's ghost*
- Figure 1.3 ..... Knowlton's virtual keyboard display*
- Figure 1.4 ..... Optical-see-through display*
- Figure 1.5 ..... Video-see-through display*
- Figure 1.6 ..... Logical structure of thesis*
- Figure 2.1 ..... Image from the ARVIKA project*
- Figure 2.2 ..... Mock up to illustrate virtual staircase in real environment*
- Figure 2.3 ..... Superimposing underlying structural elements*
- Figure 2.4 ..... Virtual fetus inside womb of pregnant patient*
- Figure 2.5 ..... HMD view simulating invasive laparoscopic procedure*
- Figure 2.6 ..... Augmented reality textual annotation of an engine*
- Figure 2.7 ..... Virtual lines show a planned motion of a robot arm*
- Figure 2.8 ..... Conceptual diagram of ARVISIA components*
- Figure 2.9 ..... Mock up images to show virtual restoration of a museum artefact*
- Figure 3.1 ..... Photograph of monitor showing a virtual cube passing under a real archway*
- Figure 3.2 ..... Real Lego™ wall with virtual brick inserted in top row; 2<sup>nd</sup> from the right*
- Figure 3.3 ..... A virtual box reflected in a real mirror*
- Figure 3.4 ..... Conceptual diagram of video-see-through augmented reality system*
- Figure 3.5 ..... A video-see-through HMD*
- Figure 3.6 ..... Conceptual diagram of optical-see-through augmented reality system*
- Figure 3.7 ..... An optical-see-through display*
- Figure 3.8 ..... Monitor-based augmented reality*



- 
- Figure 3.9 ..... *Spatially augmented reality showing scene without and with projected imagery*
- Figure 3.10 ..... *CIE Chromaticity diagram showing RGB gamut*
- Figure 4.1 ..... *Illustration to show convergence for a far and a near object*
- Figure 4.2 ..... *Illustration showing lens shape during accommodation on a far and a near object*
- Figure 4.3 ..... *A white rectangle; first unoccluded, then partially occluded by a chair*
- Figure 4.4 ..... *Illustration showing depth cueing due to shading*
- Figure 4.5 ..... *Illustration showing depth cueing due to perspective.*
- Figure 4.6 ..... *Texture gradient across field of cereal*
- Figure 4.7 ..... *The distant chess piece appears higher in the image than one that is nearer the viewer*
- Figure 4.8 ..... *Illustration showing depth cueing due to contours*
- Figure 4.9 ..... *Illustration showing depth cueing due to aerial perspective*
- Figure 5.1 ..... *Deregowski's apparatus for estimating apparent depth in pictures*
- Figure 5.2 ..... *Modified apparatus for estimating apparent depth in images*
- Figure 5.3 ..... *Apparatus dimensions*
- Figure 5.4..... *Viewing pipe with beam splitter in place*
- Figure 5.5 ..... *Real scene object design*
- Figure 5.6 ..... *Sun position*
- Figure 5.7 ..... *Test scenes*
- Figure 5.8 ..... *Left and right eye viewing effect on apparent scale alignment*
- Figure 5.9 ..... *Spectral performance of beam splitter*
- Figure 5.10 ..... *Stereo-acuity pre-test apparatus*
- Figure 5.11 ..... *Box Plot showing estimated distances for each image*
- Figure 5.12 ..... *Cast shadows influence perceived object elevation and depth*
- Figure 6.1 ..... *Display surfaces for an optical-see-through system*
- Figure 6.2 ..... *Original pixelA*
- Figure 6.3 ..... *Original pixelB*

- 
- Figure 6.4 ..... *pixelA in shadow*
- Figure 6.5 ..... *pixelB in highlight*
- Figure 6.6 ..... *pixelA shadow mask*
- Figure 6.7 ..... *pixelB shadow mask*
- Figure 6.8 ..... *pixelA overlay*
- Figure 6.9 ..... *pixelB overlay*
- Figure 6.10 ..... *pixelC original*
- Figure 6.11 ..... *pixelC in shadow*
- Figure 6.12 ..... *pixelC shadow mask*
- Figure 6.13 ..... *pixelC shadow masking colour swatch*
- Figure 6.14 ..... *pixelC in highlight*
- Figure 6.15 ..... *pixelC overlay*
- Figure 6.16 ..... *pixelC highlight overlay colour swatch*
- Figure 6.17 ..... *Real background*
- Figure 6.18 ..... *Desired composite*
- Figure 6.19 ..... *Calculated mask*
- Figure 6.20 ..... *Calculated overlay*
- Figure 6.21 ..... *Image pipeline showing ideal compositing*
- Figure 6.22 ..... *Real background with red shift*
- Figure 6.23 ..... *Image pipeline showing error in composite due to colour  
inaccuracy*
- Figure 6.24 ..... *Holdout matte*
- Figure 6.25 ..... *Overlay for use with holdout matte*
- Figure 6.26 ..... *Image pipeline incorporating holdout matte*
- Figure 7.1 ..... *Optical-see-through display employing transmissive alpha  
channel mask*
- Figure 7.2 ..... *Display prototype; plan, side, elevation and perspective  
views*
- Figure 7.3 ..... *Display prototype; exploded view*
- Figure 7.4 ..... *Display prototype with display panels in place*
- Figure 7.5 ..... *Schematic diagram of prototype system*

- 
- Figure 7.6 ..... Component models for an ideal augmented reality system*
- Figure 7.7 ..... Test scene 1*
- Figure 7.8 ..... Test scene 2*
- Figure 7.9 ..... Test scene 3*
- Figure 7.10 ..... Test scene 4*
- Figure 7.11 ..... Test scene 5*
- Figure 7.12 ..... Test scene 6*
- Figure 7.13 ..... Schematic for production of mask and overlay*
- Figure 7.14 ..... Scene 1, mask and overlay*
- Figure 7.15 ..... Scene 2, mask and overlay*
- Figure 7.16 ..... Scene 3, mask and overlay*
- Figure 7.17 ..... Scene 4, mask and overlay*
- Figure 7.18 ..... Scene 5, mask and overlay*
- Figure 7.19 ..... Scene 6, mask and overlay*
- Figure 7.20 ..... Augmented images*
- Figure 7.21 ..... Shadow casting*
- Figure 7.22 ..... Geometric model of the scene*
- Figure 7.23 ..... Superimposed reflection and transmission images*
- Figure 7.24 ..... Superimposed reflection and transmission images*
- Figure 7.25 ..... Animation sequence; masks and overlays*
- Figure 7.26 ..... Photograph through an optical-see-through display without  
masking*
- Figure 7.27 ..... Images composited with real-world view to show occlusion of  
a real background by a virtual object*
- Figure 7.28 ..... Images composited with real-world view to show a virtual  
shadow on a real object*
- Figure 7.29 ..... Optical arrangement to solve out-of-focus problem*
- Figure 7.30 ..... Projection display arrangement*
- Figure 7.31 ..... Correcting for oblique projection*
- Figure 7.32 ..... Viewing the scene through the projection prototype*
- Figure 7.33 ..... Scene with mask and with overlay*

- 
- Figure 7.34 ..... *Display arrangement using two masking projectors*
- Figure 7.35 ..... *Digital Micromirror Device and magnified view with one mirror removed*
- Figure 7.36 ..... *Left: Projection using a single DMD*  
*Right: Projection using a separate DMD for each of RGB*
- Figure 7.37 ..... *Augmented reality display arrangement using a DMD*
- Figure 7.38 ..... *Image splitting using DMD*
- Figure 8.1 ..... *3D object shown hatched with its inferred visual hull in bold*
- Figure 9.1 ..... *Lambertian reflection*
- Figure 9.2 ..... *Specular reflection*
- Figure 9.3 ..... *Specular highlight*
- Figure 9.4 ..... *Radiosity form-factor parameters*
- Figure 9.5 ..... *Hemicube*
- Figure 9.6 ..... *Element pairings between perpendicular 2D patches*
- Figure 9.7 ..... *Pairings between perpendicular 2D patches using a multi-level hierarchy*
- Figure 9.8 ..... *High dynamic range radiance maps produced by photographing probe at different exposure settings*
- Figure 9.9 ..... *Range of possible directions of light sources causing diffuse reflection at point P on a surface*
- Figure 9.10 ..... *Range of possible directions of light sources causing diffuse reflection at point P on a surface*
- Figure 9.11 ..... *Influence of a single facet on two patches may be impossible to reconcile*
- Figure 9.12 ..... *Radiosity rendering showing colour bleeding*
- Figure 9.13 ..... *Scene data structure used by rendering software*
- Figure 9.14 ..... *Platonic solids*
- Figure 9.15 ..... *Alternative sub division of triangular facets*
- Figure 9.16 ..... *Light source illuminating Surface S*
- Figure 9.17 ..... *Surface S, red radiosities and appearance*
- Figure 9.18 ..... *Faceted enclosure laid flat*

- 
- Figure 9.19* ..... *Surface S inside enclosure*
- Figure 9.20* ..... *Construction for form-factor calculation*
- Figure 9.21* ..... *Enclosure facet radiosities*
- Figure 9.22* ..... *Surface S and enclosure showing lit facets*
- Figure 9.23* ..... *Virtual rendering of surface S*
- Figure 9.24* ..... *Enclosure showing original surface S and virtual surface T*
- Figure 9.25* ..... *Resultant radiosities for surface T*
- Figure 9.26* ..... *Test scene*
- Figure 9.27* ..... *Tessellated reconstruction of test scene*
- Figure 9.28* ..... *Tessellated reconstruction of test scene with faceted enclosure shown in cyan*
- Figure 9.29* ..... *Plan view of brick and underlying floor to show obscured facets*
- Figure 9.30* ..... *Plan view of brick and underlying floor showing facets, outlined in red, created where original facet centres are obscured*
- Figure 9.31* ..... *Facets occluded by the brick. Floor facets are shaded green and wall facets yellow*
- Figure 9.32* ..... *Re-rendered scene*
- Figure 9.33* ..... *Bar plots showing original radiosities of the 16 patches of Surface S and the proportional error after re-illumination*

## **Acknowledgements**

I would like to take this opportunity to thank the following people, without whose support this work would not have been possible.

Professor Roy Kalawsky of Loughborough University for his long patience, encouragement, and sound advice.

My employers and work colleagues for allowing the necessary space, both spatial and temporal.

My sons, Andrew, Richard and David, for inspiration and encouragement.

My wife, Lynda, for proofreading, sustenance, moral support and unfailing love.

My Mother, and my Father, who died during the course of this work and would have dearly liked to see it completed.

## Abstract

In contrast to virtual reality, which immerses the user in a wholly computer-generated perceptual environment, augmented reality systems superimpose virtual entities on the user's view of the real world. This concept promises to fulfil new applications in a wide range of fields, but there are some challenging issues to be resolved. One issue relates to achieving accurate registration of virtual and real worlds. Accurate spatial registration is not only required with respect to lateral positioning, but also in depth. A limiting problem with existing optical-see-through displays, typically used for augmenting reality, is that they are incapable of displaying a full range of depth cues. Most significantly, they are unable to occlude real background and hence cannot produce interposition depth cueing. Neither are they able to modify the real-world view in the ways required to produce convincing common illumination effects such as virtual shadows across real surfaces. Also, at present, there are no wholly satisfactory ways of determining suitable common illumination models with which to determine the real-virtual light interactions necessary for producing such depth cues.

This thesis establishes that interpositioning is essential for appropriate estimation of depth in augmented realities, and that the presence of shadows provides an important refining cue. It also extends the concept of a transparency alpha-channel to allow optical-see-through systems to display appropriate depth cues. The generalised theory of the approach is described mathematically and algorithms developed to automate generation of display-surface images. Three practical physical display strategies are presented; using a transmissive mask, selective lighting using digital projection, and selective reflection using digital micromirror devices. With respect to obtaining a common illumination model, all current approaches require either prior knowledge of the light sources illuminating the real scene, or involve inserting some kind of probe into the scene with which to determine real light source position, shape, and intensity. This thesis presents an alternative approach that infers a plausible illumination from a limited view of the scene.

## **1.0 Introduction**

### **1.1 From virtual to augmented reality**

As the power-to-cost ratio for computer systems has increased substantially, there has been expanding interest in electronic graphic systems that place the human operator within a virtual environment in which as many human senses as possible are isolated from real-world experience and fed by computer stimuli. This concept of virtual reality has caught the imagination of a wider public, and research continues to devise techniques to increase realism through faster and more convincing graphics, as well as to increase the bandwidth of the virtual experience through provision of effective haptic and kinaesthetic feedback. Such systems fulfil, and promise, numerous applications in areas such as education, training, entertainment, and design visualisation. However, a fully immersive virtual reality approach suffers from a number of drawbacks.

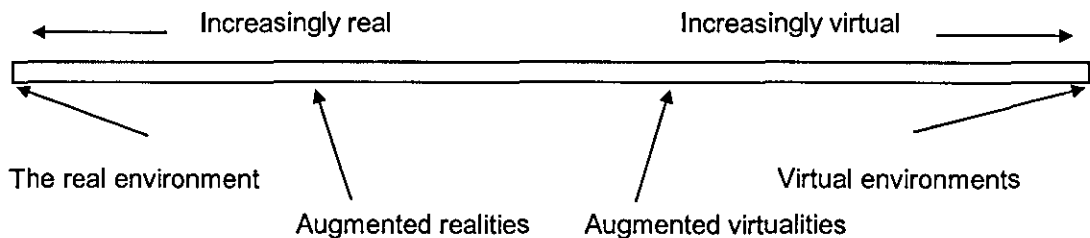
One problem, currently the subject of research, is finding ways to avoid the feelings of vertigo and nausea that can arise when using virtual reality systems [BIOC92]. Additionally, in many applications it is desirable for the virtual environment to be as realistic as possible to give it the credibility that will allow its users to suspend disbelief. This requirement drives the development of environment models that seek to better represent our real world perceptions. However, artificially simulating anything even loosely approximating a real-world environment carries a heavy computational overhead. Considerable effort is required to provide significant levels of detail in areas that may not even be the prime focus of the task, such as realistic landscapes for flight simulators and for architectural walk-throughs. Another disadvantage of immersive virtual reality systems is that users are insulated from the real world and cannot readily interact with it to carry out real tasks. For collaborative systems a further overhead may be introduced by the need to model virtual representations of the human participants.



Rather than replacing the real environment with one that is wholly artificial, the last decade has seen increasing interest in using computers to augment real experience, either through supplementing the real with the virtual, or by enhancing the virtual with the real.

The possibility of adding computer-fabricated graphics to a view of reality, or placing real objects into a virtual environment, makes it apparent that fully immersive virtual reality systems actually lie at one end of a continuum of what have been described as mixed reality systems [MILG94] (figure 1.1). At the other extreme of this continuum lies reality itself, and classified between these poles are a range of systems that mix, to varying degrees, real objects with those that are computer generated.

*The mixed reality continuum:*



(Figure 1.1 – The Mixed Reality Continuum)

These hybrid systems may mix virtual elements with reality to provide a predominantly synthetic view that incorporates some real objects, or they may supplement the user's experience of a real environment by the addition of computer stimuli. The term augmented reality has been coined to describe systems falling into the latter category. Such computer enhancement of reality offers advantages over virtual reality by not only potentially avoiding the need for complex modelling of people and environment, but also by providing an anchor in reality that should reduce the likelihood of nausea being induced. In addition, augmented reality

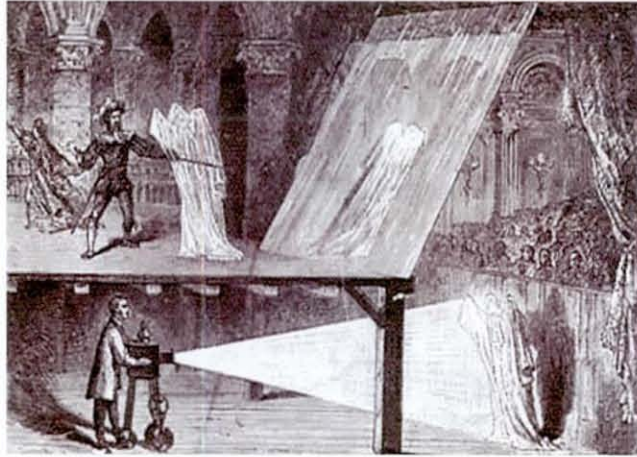
systems promise to allow the operator to actively carry out tasks involving real-world objects rather than being confined to a wholly artificial environment.

## **1.2 Suspending disbelief**

A broad distinction can be made between systems that are intended to mix the real and virtual such that the viewer is led to believe that the synthetic elements are part of reality, and those systems that overlay information on reality with no intention of leading viewers to believe that the virtual entities exist in the real world. An example of the first would be the way in which filmmakers mix computer animation and live action with the intention of suspending audience disbelief. The latter is exemplified by modern SLR (Single Lens Reflex) cameras, which show exposure data superimposed via the viewfinder optics. In this thesis, the emphasis is on systems that seek to integrate computer graphics with the user's view of their actual real-world environment.

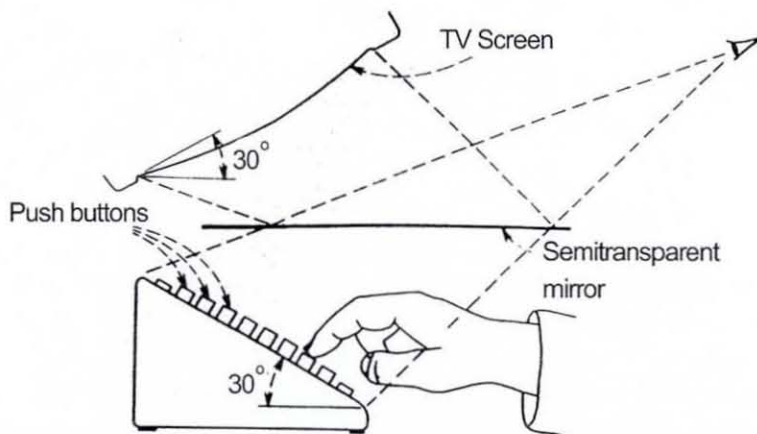
## **1.3 Historical background**

The historical roots of mixing the real and the virtual with the intention of making viewers believe that they are one and the same can be traced back at least to the 1860s when John Henry Pepper and Henry Dircks applied for a patent describing a system for producing theatrical effects [WALK94]. In its early incarnation, this simply involved placing a large pane of plate glass set at a 45° angle on the stage. An object behind the glass could be lit and thus made visible to the audience or a well-lit off-stage object could be made to appear on stage by reflection in the glass. By carefully controlling the relative illumination, one object could be made to 'morph' visually into the other. In figure 1.2 the ghost actor is located below and forward of the stage and shown on the left is what the audience see. The effect was first demonstrated at the London Polytechnic Institute where Pepper was a professor. The 'Pepper's Ghost' technique is still used in theatres and can be found in some amusement park attractions such as Disney's Haunted Mansion and the Haunted House at Alton Towers in the UK.



(Figure 1.2 – Pepper's ghost)

Much the same optical arrangement, but on a smaller scale, was later used by Knowlton at AT&T [KNOW77] when he developed a system that optically superimposed computer displays onto an input device. The purpose was to allow users to interact with a physical keyboard, whilst also providing flexibility of function by optically superimposing alternative labels onto the keys. In this way, the same real keyboard could be given the appearance of a typewriter, a calculator or a telephone operator's console. In Knowlton's system, shown in figure 1.3, the computer graphics were seen reflected in a semi-transparent mirror that was positioned over the keyboard at an appropriate angle, in a Pepper's-Ghost-type optical arrangement.



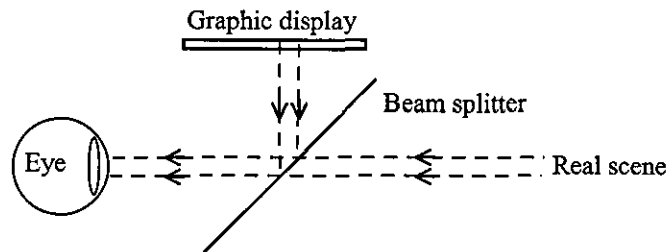
(Figure 1.3 – Knowlton's virtual keyboard display)

The early development of wearable augmented reality displays was pioneered by Sutherland [SUTH68] and found impetus in head-mounted systems for fighter pilot and military helicopter situation displays [FURN69]. These displays were similar to those now used for immersive virtual reality but, instead of presenting only video or graphics to the user, the optical elements allow the user to see the graphics superimposed on the view of the real world.

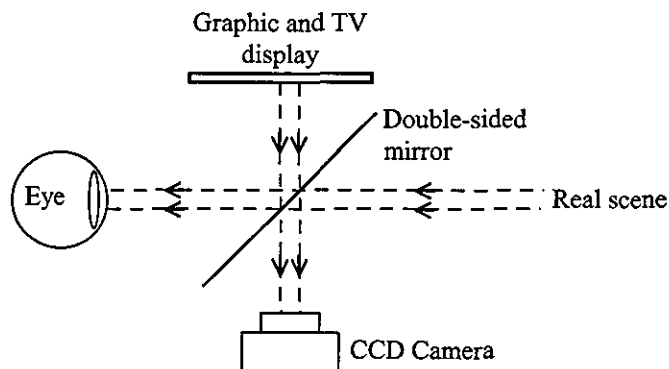
Different techniques for mixing real and virtual have and are being explored, often for the purpose of artistic expression. One such early exploration is provided by the work of Krueger who, in 1974, coined the term '*artificial reality*', and developed a system called GLOWFLOW [KRUE91]. GLOWFLOW provided an environment that had the capability, albeit limited, of responding to people within it. The environment space was a darkened rectangular gallery with walls containing transparent tubes, which in turn contained coloured phosphorescent particles suspended in water. Hidden light sources activated the phosphors to give the illusion of objects floating in space. Sound was generated using a Moog synthesiser, and the whole system controlled by a PDP-12 minicomputer so that, when viewers moved around on pressure-sensitive floor pads, the sounds and optical effects altered. Krueger's work on artificial reality continued with the development of VIDEOPLACE [KRUE85] in which users' active video images, portrayed in an artificial world, could interact with graphic objects and with one another. Krueger later implemented VIDEODESK [KRUE91] which utilised a light table with a camera mounted above, so that the silhouette of a user's hands could be superimposed on a computer application and the fingers used to point, draw, or edit screen graphics and text. Most significantly, Krueger sought to free the interactor from the encumbrance of headset and gloves through the use of video technology and gesture recognition, and to integrate virtual worlds with reality by effectively placing the user within the virtual environment. Other research, in the same spirit, includes Baudel and Beaudouin-Lafon's [BAUD93] work on Charade; a system allowing the remote control of objects using freehand gestures, and Helsinki University's Virtual Orchestra [HIIP97] demonstrated at SIGGRAPH '97.

#### 1.4 Optical-see-through versus video-see-through

Pepper's Ghost, Knowlton's system and pilot head-mounted displays all use optical elements that allow the user to see the virtual entities superimposed on a direct view of reality and this approach is usually termed 'optical-see-through' augmented reality. (This display approach is shown diagrammatically in figure 1.4.) However, graphics can also be visually added to reality by superimposing computer imagery on a video image of a real scene. This principle is common in the television industry where subtitles and other graphics are combined with live video. In similar fashion, non-see-through, head-mounted displays can be used to provide images generated by mixing computer graphics and video camera input of the user's real environment. Such displays are often referred to as 'video-see-through', as they superimpose graphics on a video view of reality. (See figure 1.5.)



(Figure 1.4 – Optical-see-through display)



(Figure 1.5 – Video-see-through display)

### **1.5 Motivation**

Although we are only just beginning to see fully tractable applications for augmented reality, its future offers significant promise. It affords the possibility of enhancing a user's senses providing; additional information about their environment, X-ray vision, low-vision aids, new forms of entertainment, guidance for construction and maintenance tasks, design visualisation and help with remote collaborative working. For example, doctors could use augmented reality to view underlying organs or tumours for surgery. Patients suffering from conditions such as macular degeneration could wear augmented reality spectacles that would enhance the retinal image in localised areas to counter loss of vision. Maintenance engineers could see instructions superimposed on the machine upon which they are working. Service engineers could don glasses enabling them to see hidden wiring or ducting as they tour a building. Fire fighters could wear visors, which would display routes through buildings despite dense smoke. Visitors to archaeological sites could see ruins apparently reconstructed to their former glory or architects judge the impact of a proposed building on the actual site. Interior designers could try-out colour schemes and furnishings in a virtual way but within a real space. These, and other possibilities are discussed at greater length in Chapter 2.

### **1.6 Implementation hurdles**

As outlined above, augmenting reality offers a wide range of application possibilities. However a number of technical problems need to be resolved if its promise is to reach fruition. These issues, in the main, relate to parameters that influence the extent to which convincing integration of graphics and reality can be achieved dynamically. Accurate registration of real and virtual entities is of particular importance as the degree of tolerance to error in the position of a graphic overlay relative to the real world is very small for augmented reality systems. This puts particular demands on the accuracy of tracking systems and the speed at which image update can be achieved when the user's view changes. For some applications, the suspension of disbelief may also be of significant importance. This means that the virtual elements must appear to belong in the real scene as convincingly as

possible. In these cases, the virtual elements must be generated in ways that not only register properly but also correctly exploit depth perception cues so that synthetic entities appear integrated appropriately within the real 3D world. Although there is a long and substantial heritage of research literature concerned with aspects of human perception in the real world and, more recently, issues relating to perception in virtual realities have begun to be explored, very little consideration has been given to the perceptual effects of mixing the virtual with the real. Additional problems arise due to the inability of existing see-through augmented reality processing and displays to render synthetic graphics in a way that is suitably integrated with reality.

### **1.7 Thesis statement**

The aim of the work described in this thesis is to investigate ways of improving visual quality when integrating virtual entities with real scenes using optical-see-through displays. A key issue appears to be the fact that current optical-see-through systems are unable to simulate occlusions and shadows appropriately. These provide significant depth cues in the real world, but their relative importance has never been quantified for augmented reality environments. Unfortunately, the optical arrangement on which displays are based offers no possibility of producing these cues, as it is unable to modify the real-world view in the ways that would be necessary. Also, there are no currently available techniques for acquiring the common illumination models needed to produce realistic real-virtual shading and shadow interactions without knowledge of all light sources affecting the real scene.

In order to address these problems, this thesis details investigations carried out:

- to determine the relative importance of interposition and shadow in the judgement of depth in augmented realities,
- to find a way of displaying mutual real-virtual occlusions and shadows while maintaining a direct view of the real world,
- to develop a technique for acquiring common illumination models from real-world views.

These investigations lead to the theses that:

- the realistic visual integration of virtual entities in a real environment depends on appropriate display of occlusion and shadow effects that cannot be simulated in existing optical-see-through augmented reality systems;
- current limitations with respect to producing occlusion and shadows, while maintaining a direct view of the real world, can be overcome by incorporating a real-world mask into optical-see-through displays;
- it is possible to deduce a plausible real-world illumination model from local scene knowledge alone; something that cannot be achieved currently.

To defend these theses I make the following contributions.

- Design and execution of an experiment to show that the presence of cast shadow can provide a more accurate impression of depth, in a static composite display, than can be obtained from interposition alone.
- An extension of the alpha-channel image transparency concept to viewing of the real world so that interposition and shading can be realised in optical-see-through augmented reality displays.
- Design, construction and demonstration of prototype optical see-through displays to implement a real-world alpha-channel.
- Conceptual design of an improved augmented reality display using digital micro-mirror devices to implement a real-world alpha channel
- Development and proof-of-concept demonstration of a strategy for inferring a plausible common global illumination model.

The work begins with an experiment that quantifies the relative significance of interposition and cast shadow in providing appropriate static depth cueing for augmented reality. The results are used to justify what, perhaps, already seems intuitive; that is the importance of establishing common geometric and illumination models between reality and any superimposed synthetic entities.



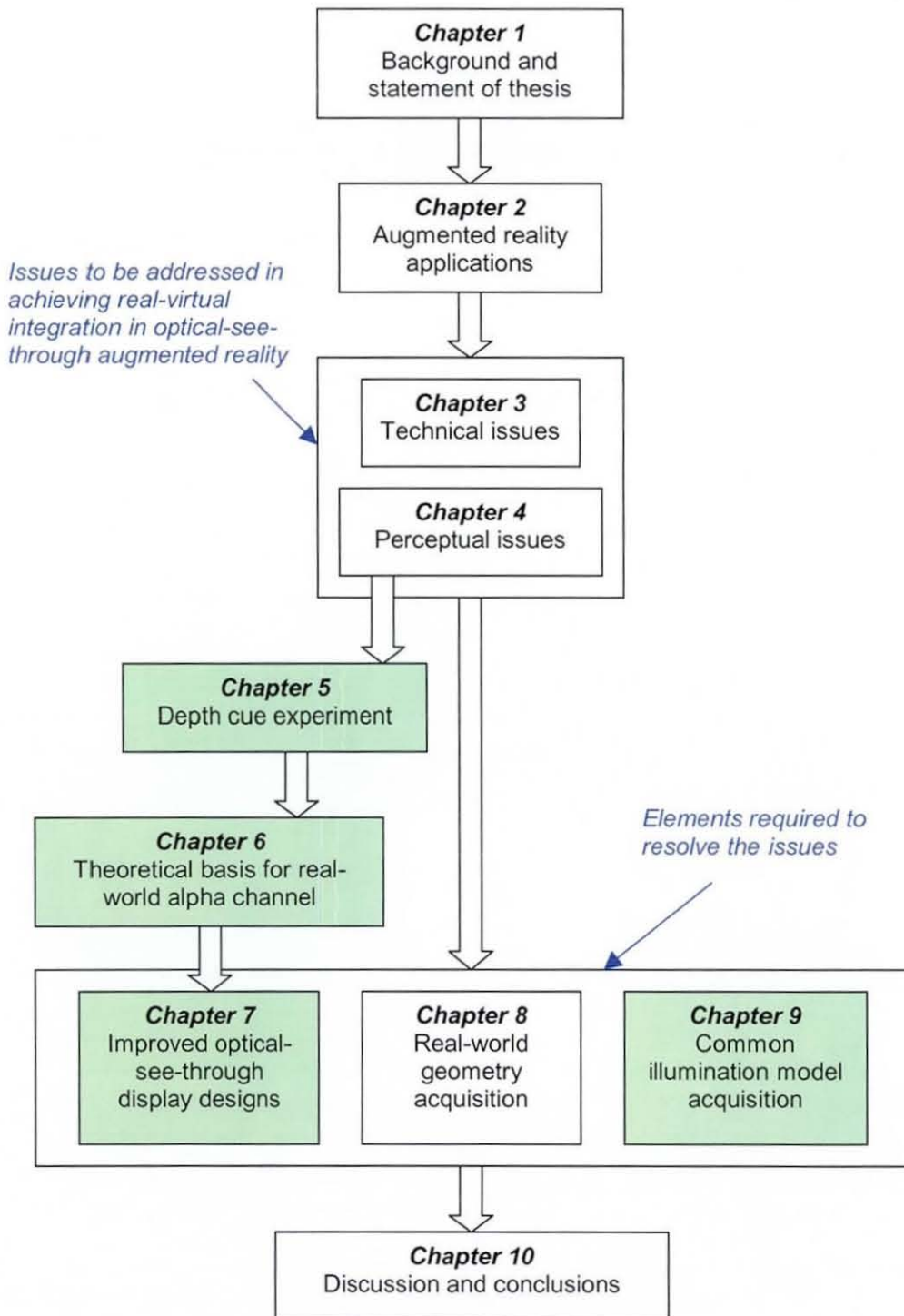
The extent to which convincing visual integration depends on interposition and illumination effects exposes a fundamental limitation of existing optical-see-through displays, in particular their inherent lack of control over transparency. To address this problem, two novel display arrangements are presented and demonstrated, and a further design proposed. Acquiring a suitable common illumination model presents a particularly difficult challenge when the real-world view is not a wholly enclosed environment with all light sources known *a priori*. To address this situation, at least in part, a radiosity inference model is developed and verified empirically for simple scenes.

Although it is the case that computers could play a role in the synthetic augmentation of information reaching any of our human senses, this thesis considers only the visual channel. However, other possibilities are discussed in chapter 10.

Motivation for the work is established through a survey of the range of possible applications of augmented reality presented in Chapter 2. Chapters 3 and 4 discuss the technical and perceptual issues to be addressed if the envisaged applications are to be fully realised. The importance of interposition and shadow depth cues are quantified by an experiment described in Chapter 5. Chapters 6 and 7 develop the theoretical and practical basis for compositing scenes using real-world alpha-channel masking.

Prerequisite for full real-virtual compositing is the ability to acquire a model of reality. The geometric aspects of this are considered in Chapter 8, while in Chapter 9 the issue of common illumination is discussed and a new approach presented. The thesis is concluded in Chapter 10.

Figure 1.6 illustrates the logical structure of the thesis indicating the content of each chapter. The shaded boxes represent the areas in which this thesis makes novel contribution.



(Figure 1.6 – Logical structure of thesis)

## **2.0 Potential applications of augmented reality**

The development of fully tractable computer-based augmented reality systems is in its infancy, with almost all substantive progress having been made within the last decade. A wide range of potential applications exists, and includes; assistance for manufacturing and maintenance, construction, medical imaging, annotating the real world, training, teleoperation of robots, design visualisation and collaborative working. This chapter reviews the possibilities for the application of augmented reality that provide the motivation for this thesis.

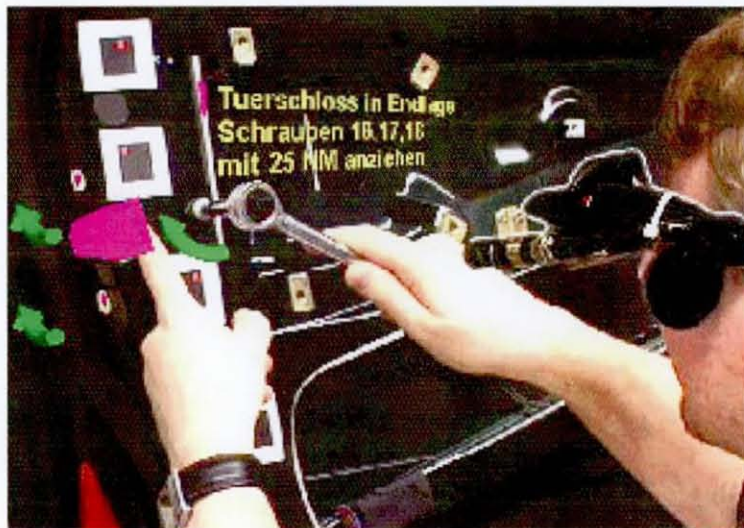
### **2.1 Manufacturing and maintenance**

Augmented reality systems have potential for aiding manufacturing and maintenance. For example, graphics superimposed on a view of a real object could be used to show cutting or drilling positions, or could provide assembly guidance. Similarly, it would be possible, equipped with suitable see-through headgear, to be given step-by-step instructions while actually working on the repair of a system.

Along these lines, Caudell and Mizell [CAUD92], of Boeing Computer Services, point to the impracticality of automating the manufacture and assembly of aircraft due to the small average lot size for many of the parts and the high dexterity required for some aspects of assembly. Also, with the greater complexity of aircraft, assembly engineers are required to use an increasing amount of information in the form of assembly guides, templates, wiring lists, etc., and even small design changes can result in expensive delays. In a bid to overcome these problems, they propose an augmented reality system with which a worker, equipped with see-through virtual reality goggles, can view his work piece with appropriate position markers and other assembly information graphically superimposed.

Researchers at Columbia University have been experimenting with a similar use of augmented reality to annotate real world objects and thus aid the performance of 3D tasks. Their test-bed system, called KARMA (Knowledge-based Augmented Reality for Maintenance Assistance) [FEIN93] uses augmented reality to explain laser printer maintenance and repair tasks by the knowledge-based generation of graphics to overlay the user's view of the physical world. By visually superimposing simple 3D line graphics on the user's view of a real printer, the system provides guidance for carrying out basic tasks such as removing the paper tray.

The ARVIKA project [ARVI03] is probably the largest current augmented reality research effort focusing on engineering and maintenance. This project aims to develop augmented reality systems that are able to support work processes in development, production and servicing in engineering by visually superimposing instructions to support real-world tasks, as illustrated in figure 2.1.



(Figure 2.1 – Image from the ARVIKA project)

## 2.2 Construction

There is also potential for augmented reality to be applied to architecture, civil engineering and construction. For example, the technology promises to provide direct visualisation of proposed new structures on an existing site, and within a real landscape, allowing assessment of visual impact without the need for pre-determined choreography or detailed graphic modelling of neighbouring structures or natural elements such as pre-existing trees. It also makes feasible, visualisation of the anticipated results of remodelling a landscape, structure or interior by visual replacement of existing elements with computer-generated alternatives, whilst retaining a direct view of the unchanged aspects of the real environment.

To illustrate the visualisation of new structure within an existing building, figure 2.2, shows two views of a virtual staircase superimposed on a real scene, mocked-up using paint-package software.

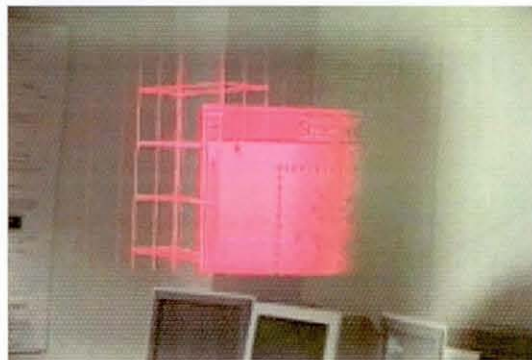


*(Figure 2.2 – Mock-up to illustrate virtual staircase in real environment)*



Overlaying visual paths and instructions on reality also promises to provide an unambiguous and rapidly modifiable aid to the construction process and to building maintenance. Superimposition of computer-generated graphics over a direct, real view of a structure can allow comparison of an 'as-built' state against the intended design model. In addition, it becomes possible for on-site visualisation of invisible objects or effects such as hidden structural elements or heat flows to be realised. Being able to see the exact location of wiring or piping within a wall would help avoid damage and aid planning for remodelling.

Feiner et al. at Columbia University [FEIN95] have conducted some early work in this area. Their system uses a see-through head-mounted display to overlay a graphical representation of a building's structural elements on the user's view within the building (figure 2.3). The same team has also applied augmented reality to aid the construction of 3D building units called space frames [WEBS96].



(Figure 2.3 – Superimposing underlying structural elements, University of Columbia)

### 2.3 Medical imaging

Augmented reality also offers exciting possibilities for medical imaging. It could be used to aid navigation during difficult procedures or to enable a surgeon to see underlying anatomical structures visually superimposed on the patient's body. For example, a team at the University of North Carolina (UNC) [BAJU92] have developed an experimental system to enable ultrasound scans to be viewed superimposed in position over the abdomen of a pregnant woman (figure 2.4), using a video-see-through head-mounted display, with a head-mounted camera providing video input from the real scene.



*(Figure 2.4 - Virtual foetus inside womb of pregnant patient, UNC Chapel Hill Dept. of Computer Science)*

This concept has been further developed at UNC to aid procedures that currently require difficult hand-eye coordination and 3D visualisation skills such as needle biopsies [FUCH96] and laparoscopic surgery [FUCH98]. (See figure 2.5.)



*(Figure 2.5 - The HMD view simulating a minimally invasive laparoscopic procedure, UNC Chapel Hill Dept. of Computer Science)*

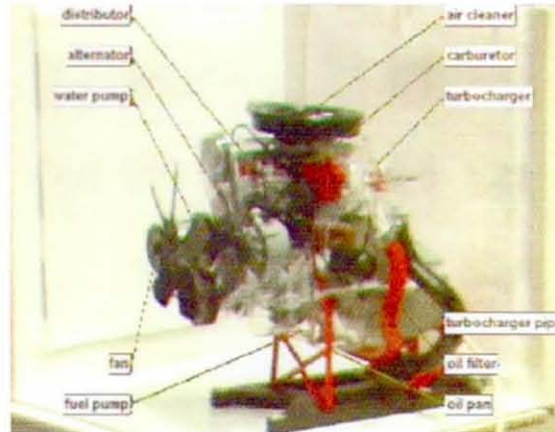
In similar vein, the ARTMA Virtual Patient™ System [GUNK95] uses augmented reality to aid navigation in endoscopic surgery. During an operation the surgeon watches the live endoscopic video on a monitor; the region of interest and the desired path of the endoscope are superimposed graphically. In effect, the surgeon has to follow a series of rectangles floating in space marking the trajectory of best approach. Artma Medical Technologies Inc. of Salt Lake City are also developing a system to allow a dentist to use a head-mounted display to see the drilling position relative to a planned dental implant [WATZ99].

With regard to using augmented reality for surgical guidance, the Computational Imaging Science Group at the Department of Radiological Sciences of Guy's Hospital, London, are developing a system called MAGI which superimposes computer-generated images in stereo onto the view through an operating microscope. The purpose is to enable surgeons to view hidden critical structures such as tumours and arteries [EDWA03].

#### **2.4 Annotating the real world**

Augmented reality could also provide ready access to other information relating to real-world objects and environments. For example, Rose et al. [ROSE95] have been developing an augmented reality system for annotating real-world objects. In their application they superimpose graphic labels on the view of a real car engine. The user interacts with the real object in a natural environment and a position-sensed pointing device is used to point at its parts. Appropriate annotations are superimposed on a video display of the engine, which is tracked so that annotations move if the view orientation is changed (figure 2.6).





(Figure 2.6 – Augmented reality textual annotation of an engine from the User Interaction and Visualization Group at the former European Computer Industry Research Centre (ECRC), Munich)

Another area of application is to allow the visualisation of data from sensing devices. These could be measuring properties of objects or environment such as; temperature, pressure, radioactivity, voltage, or gas emission. For example, researchers at Georgia Institute of Technology [OCKE98] have developed a wearable system to facilitate the collection of temperature probe data in a chicken processing plant. The aim is to solve the problem of obtaining real-time quality assurance data using highly mobile plant workers in a very noisy environment. Although the system uses a head-mounted display, no position tracking or real-world visual registration is employed.

A further example of using augmented reality to annotate the real world is provided by the '*Touring Machine*' [HOLL99] developed at Columbia University. This system uses a head-mounted display to superimpose information about the university campus as the user wanders through it. Tracking is achieved using GPS (Global Positioning System) data, with magnetic compass and pitch and roll transducers to determine head orientation.

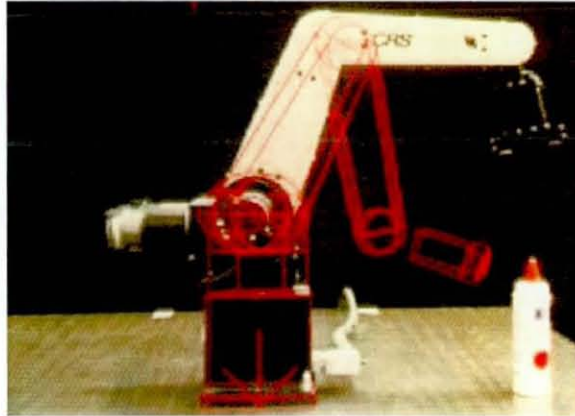
## 2.5 Education and training

Augmented reality systems could also become the vehicle for innovative and flexible training systems. McManners, [MCMA97] Defence Correspondence for the *Sunday Times* reports on CATT (Combined Arms Tactical Trainer); a system which 'arms' soldiers with virtual reality helmets and eventually should allow battle participants to be graphically superimposed on real terrain. Similarly, Metzger, [METZ93] of Loral Advanced Distributed Simulation Inc., describes a military training system capable of overlaying images of the real world onto a virtual scene, as well as overlaying virtual objects onto a real scene. A soldier equipped with adapted binoculars or head-mounted display sees virtual elements superimposed on the real landscape. For example, one envisaged scenario involves a simulated helicopter emerging from behind a real hill.

More benign training applications are also possible. For example, a virtual pair of hands superimposed on a pupil's view of a piano keyboard could, perhaps, be used to teach instrumental skills. Monitoring the user's hands could allow the system to provide context sensitive remedial help as and when required.

## 2.6 Telerobotics

Another promising application area for augmented reality lies in the remote operation of robots. Teleoperation of a robot can be a difficult problem, sometimes with long delays in the communication link. Using augmented reality, the user can plan and specify the robot's actions by real-time manipulation of the local virtual version, with the outcome directly displayed on the real world. When satisfied with the plan, the user can instruct the real robot to carry it out. Drascic and Milgram's ARGOS system has demonstrated the concept of using augmented reality as an accurate way of robot path planning [DRAS93] [MILG93]. Measurements can be made remotely and trajectory paths graphically defined [MILG95].



(Figure 2.7 - Virtual lines show a planned motion of a robot arm from David Drascic and Paul Milgram, University of Toronto)

## 2.7 Low-vision aids

There is potential for augmented reality to be used in the development of low-vision aids. In collaboration with a partner from the Paybody Eye Unit of Coventry and Warwickshire University Hospitals NHS Trust, the author has proposed a design for a low-vision aid for sufferers of Macular Degeneration. Age-related macular degeneration (AMD) is a primary cause of reduced vision in older people and such patients might suffer reduced acuity, central visual field defects and a variety of colour vision defects. In many cases the AMD condition cannot be treated. This proposal, with the cooperation of other European partners, forms the basis of the ARVISIA (Augmented Reality for the VISually ImpAired) project, which is the subject of a recent funding bid. The system would augment the user's view of the real world with computer-generated enhancement to counter an individual patient's sight impairment in a way that is localised across the retina. Such a system would also be able to correct for other common conditions such as presbyopia and colour blindness.

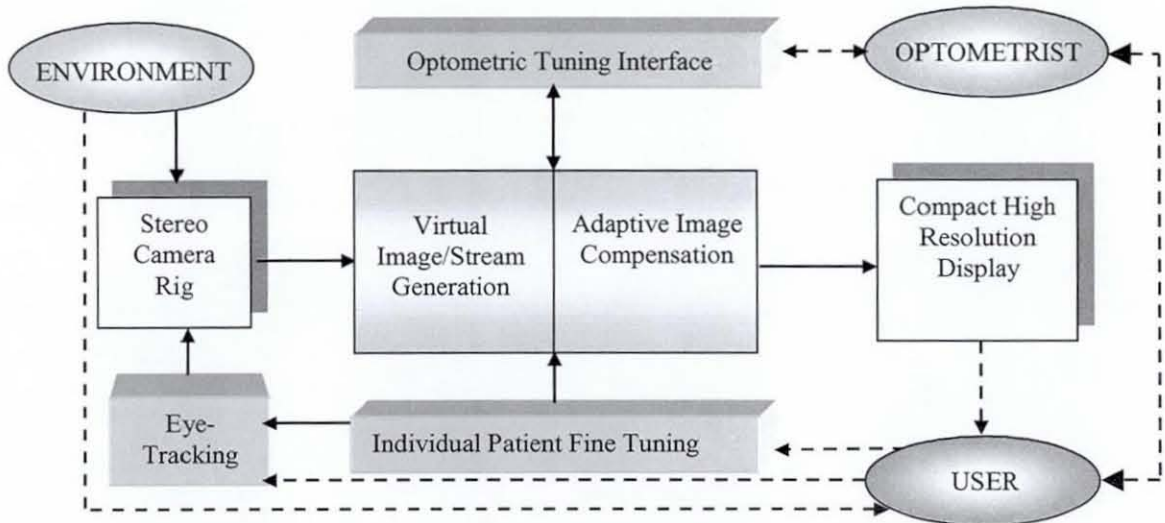
The main features of the ARVISIA project include:

- Design and construction of a non-portable vision deficiency aid using mixed and augmented reality technology



- Design and implementation of algorithms for vision capture and enhancement to counter, according to individual patient prescription, the vision deficiencies brought about by aged-related Macular Degeneration.
- Design, conduct and analysis of clinical trials evaluating the non-portable system and the prototype effectiveness for mitigating the effects of age-related vision deficiencies.
- Design, construction and evaluation of a wearable prototype vision aid.

Figure 2.8 shows a conceptual diagram of the system components.



(Figure 2.8 – Conceptual diagram of ARVISIA components)

## 2.8 Design, visualisation and collaborative working

Augmented reality also offers the potential for creative visualisation and collaborative working. Superimposing a computer representation on the user's view of the real landscape could be used to assess the visual impact of a new road or building. Fashion designers could graphically clothe human models, and interior designers experiment with superimposed wall-coverings and furniture within real

rooms. Collaborative workers could interact with virtual design objects, while maintaining real interaction with each other.

Ahlers et al. [AHLE95] pursue some of these ideas in a distributed augmented reality system for collaborative design. In their application, the objective is to allow users at remote sites to collaborate on the furniture layout within a room. Pieces of furniture represented using rendered computer graphics can be called up from an electronic catalogue, and placed within the video image of the real room. Changes made by one person can be seen on the monitors of all participants.

In similar vein, Sony is experimenting with what they call their TransVision system [REKI96], which is also designed to use augmented reality to assist collaborative design. Rather than a head-mounted display the system uses palm-top video-see-through displays, held by each participant, with which the users see a computer generated 3D image superimposed on the real-world view. The system tracks the position and orientation of one of the palm-top displays to which is attached a small CCD camera. Participants thus share the same view and may each, in turn, manipulate the 3D model. It is envisaged that such a system could facilitate designers or engineers who currently may need to build physical models to visualise and discuss a design concept. More recently, Ohshima et al. [OHSH98] have demonstrated an augmented reality collaborative system that allows two users to play air hockey using a virtual puck, while Reitmayr et al. [REIT01] allow players to participate in a game of augmented reality chess.

A further possibility for augmented reality is the virtual reconstruction of archaeological remains and museum artefacts, which could be viewed with virtual restoration superimposed. Figure 2.9 shows mock-up views of a restored ceramic plate to illustrate the concept. The author's work in this domain in collaboration with the Royal College of Art is described in Chapter 7 of this thesis.



(Figure 2.9 – Mock-up images to show virtual restoration of a museum artefact)

## 2.9 Concluding remarks

The potential applications of augmented reality are wide ranging and benefits in fields such as; medicine, design, engineering, construction, education and archaeology appear promising.

The applications of augmented reality capitalise on its ability to provide:

- annotation of the real world,
- enhancement of the real world view,
- imaging of entities that are hidden or obscured by real objects,
- positional, sequential and trajectory path guidance, and
- design visualisation.

However, the technology on which successful application depends raises many outstanding issues and, as yet, few applications are without considerable difficulties preventing commercial use. The following two chapters review these issues from technological and perceptual points of view.



### 3.0 Technical issues in implementing augmented reality

The development of systems to fulfil some of the potential applications of augmented reality is the subject of current research, with several unresolved issues outstanding. Even though computer graphics and live action are famously combined in films such as *'Who Framed Roger Rabbit'* [ZEME88] and *'Terminator 2: Judgement Day'* [CAME91] these are not achieved under the real-time, without-human-intervention requirements demanded by many proposed augmented reality applications. Significant problems still need to be solved if the proposed applications outlined in chapter 2 are to be truly viable. In relation to the augmented scenes presented to a viewer, the main technological issues can be categorised as relating to registration between real and virtual entities, the interactions between them, and the hardware display arrangement needed to present a composite image to the user. These issues are reviewed in this chapter.

#### 3.1 Registration and tracking

Registering computer graphics accurately onto a real scene presents a significant challenge requiring resolution of a number of factors to minimise static and temporal errors.

Static registration errors occur when the reconciliation of real-world, tracker, and virtual-world co-ordinate system uses inaccurate transformation values, and are also caused by the inherent inaccuracies of current tracking devices. Inaccuracies can occur in relation to tracking the observer or tracking real-world objects. Such errors may be evident even when all objects, and the user, remain stationary. Less tractable are the temporal registration errors that can occur due to system lag when an object in the world or the viewer moves. Nevertheless, real-time registration of graphics remains crucial to the success of most applications, especially those providing manufacturing or surgical guidance where positional accuracy may be vital. For accurate registration, not only must the virtual world be orthoscopic (that is, exactly

oriented and scaled with the real), requiring accurate calibration of all system components, but the location and orientation of the viewer must be continuously and precisely monitored, with the graphics almost instantly updated. These requirements present difficult challenges and have been the subject of much of the early research effort in augmented reality [AZUM93][KALA98].

The tracking requirements for augmented reality can be demanding in terms of:

- accuracy,
- latency,
- range.

Whereas small tracking inaccuracies may not be noticeable in an immersive virtual reality system, even very small angular errors in detecting the orientation of an augmented reality headset can result in a large displacement in registration of graphics with real objects that are some distance away. For example, an error tolerance of  $\pm 0.5^\circ$  of arc in measuring head orientation could result in over 17mm image displacement at 1 metre distance.

The combined latency of the tracker and the graphic system must be very low. Ideally there should be virtually no delay between change of view of the real scene and the corresponding computer graphic update. The total potential delay is caused by the time it takes for the tracker subsystem to take its measurements, plus the time it takes the corresponding images to appear on the display devices. If this combined latency is, say, 100ms (typical of many current head-mounted display based systems), a moderate head movement of  $45^\circ$  per second would produce an angular error of  $4.5^\circ$ . At faster head movements the angular error obviously increases proportionately.

Unlike tracking for virtual reality, where the range of user movement may be limited to head and upper body, tracking for augmented reality must generally operate over



longer distances. Many augmented-reality applications require that the user move about the environment, which could be within the confines of a room or, for example in the case of architectural augmented reality, could be over a large outside area.

Unfortunately, no existing methods of tracking totally fulfil these requirements. Systems with sufficient range (such as GPS; Global Positioning System) have insufficient accuracy. Conversely, more accurate tracking methods lack the necessary range.

The following sections outline the principal approaches to tracking and their relative merits with respect to augmented reality.

Tracking technologies can be based on:

- mechanical linkage,
- magnetic field sensing,
- optical sensing,
- acoustic time of flight,
- inertial transducers,
- GPS (Global Positioning System),
- hybrid approaches.

### **3.1.1 Mechanical linkage**

Mechanical tracking systems use physical linkages between the reference position and the target object. Typically, potentiometers detect the linkage angles and hence the position of the referenced object can be determined. [JAU91]. While six degrees of freedom are possible and precision is good, typically only a limited range of motions is feasible because of the kinematics of the joints and the length of each link.

### 3.1.2 *Magnetic field sensing*

In magnetic tracking devices, the relative position and orientation of receiver relative to transmitter coils are detected due to the interaction of their magnetic fields. [POTT67]. Examples of magnetic tracker include; Fastrack, Isostrack, Insidetrack and Ultratrack from Polhemus [POLH03]. This type of tracking is popular for virtual reality applications, having the advantages of; no occlusion problems, high update rate, low lag, small size and relative inexpensiveness. However, they suffer from having a low operating range, as well as a sensitivity to electromagnetic noise and metallic objects causing distortion of the magnetic field.

### 3.1.3 *Optical sensing*

Approaches to optical tracking can be categorised as those that use passive optical sensors to track a target pattern on the moving object (so-called '*outside-in*' configuration as used by Gennery [GENN92] and ARToolkit [HITL03]), and those where the sensor is on the moving object with the tracked patterns fixed in the environment. (This is called '*inside-out*' configuration and is used, for example, by the HiBall tracker system from UNC at Chapel Hill [WELC99].) '*Outside-in*' techniques are sometimes called '*vision-based*' tracking in that camera pose is estimated based on camera-image information alone. As such, these methods can require minimal specialist hardware, although lag can be significant depending on the processing that needs to be done. Unfortunately, all optical approaches are sensitive to spurious light and to occluding objects blocking line of sight between sensor and target.

### 3.1.4 *Acoustic time of flight*

Typical acoustic trackers are ultrasonic and involve three or more ultrasonic emitters on the target with three or more receivers on the reference [eg LOGI91]. Position and orientation is tracked by analysing time of flight between sensors. Acoustic trackers are generally compact in size and do not suffer from the distortions associated with some other methods. However, they have low update rate and are sensitive to occlusion and noise.

### 3.1.5 *Inertial transducers*

The principle of inertial sensing is based on the attempt to conserve either a given axis of rotation, as in the case of a mechanical gyroscope, or a position, as in the case of an accelerometer [AZUM95]. An example of a tracker of this type is GyroMouse™ from Gyration [GYRA03]. This approach has the advantages of high update rate, long range, and the fact that no reference is needed. However, errors are cumulative since each measurement is relative to the previous.

### 3.1.6 *GPS*

The GPS (Global Positioning System) tracking principle uses satellites and ground stations spread around the world to enable receivers to detect their position [ELLI96]. The resolution accomplished with such systems is usually of the order of a few metres. However, a more precise system, the differential GPS, uses emitting ground stations that refine the resolution to the order of one metre [NOE94]. A clear advantage of GPS is its range but drawbacks include poor accuracy and resolution, and the failure of the technology if the direct lines of sight to the satellites are not maintained.

### 3.1.7 *Hybrid approaches*

For many augmented reality applications, faster, more accurate methods of position and orientation tracking are required, as well as effective methods of tracking over larger distances. Whereas mechanical trackers have good accuracy, they impose constraints on motion. Magnetic tracking is limited by range and suffers potentially large errors caused by magnetic field distortions. Acoustic methods are also sensitive to noise and become less practical with increasing range. GPS is wide-ranging but not yet sufficiently accurate for most augmented reality applications. It is also subject to occlusions. Likewise, optical trackers are affected by occlusions and the more accurate ‘inside-out’ configurations require a specially built environment. Inertial trackers can decrease system lag and are not limited in range but suffer from cumulative drift errors. Hybrid approaches attempt to exploit the combined advantages of a selection of methods.



In an early attempt at combining tracking methods, Azuma added inertial tracking to an existing optical system [AZUM95] to predict user head movement. He found that, on average, prediction with head-mounted inertial sensors gave accuracies 2 to 3 times greater than prediction without inertial sensors, and 5 to 10 times better than not doing any prediction at all. Thus, future tracking for augmented may rely on similar techniques that seek to combine the respective advantages of different technologies in hybrid approaches [YOU99].

### 3.2 Interactions

In addition to the issues of registration, in order to arrive at a unified composite image, the physical interactions between real and virtual objects need to be simulated. Interactions to be considered here include:

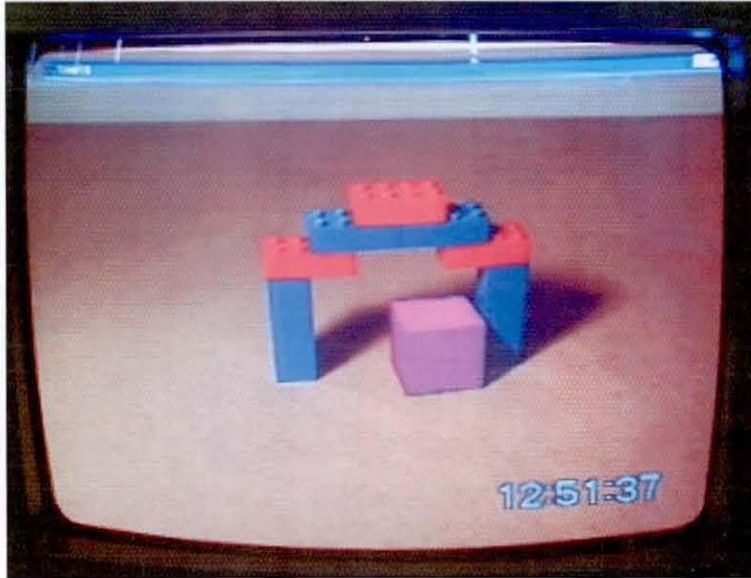
- collisions between real and virtual objects, and
- light interactions between the real and virtual.

#### 3.2.1 Collisions

Real-world objects are subject to the physical effects of gravity, friction and collision so, if convincing dynamic integration is to be achieved, virtual objects will need to appear subject to the same constraints and interactions. A virtual object placed on a real cushion should produce an appropriate visual depression in the fabric of the cushion or virtual clothing on a real person should drape and flow accordingly. Aliaga's [ALIA97] work begins to explore related issues with the development of a system that allows virtual balls apparently to bounce down a real staircase. The system relies on the prior creation of an accurate computer model of the real stairs, and the utilisation of a fast enough collision detection algorithm for an acceptable frame rate of at least 12 frames per second.

Some initial work in this area by the author is illustrated in figure 3.1, which is a frame from a video sequence of a real scene containing a Lego™ arch. The scene also contains a computer-generated cube that interacts with an invisible computer model representing the geometry of the real-world environment. The computer-

based model handles real-virtual collision detection so the cube can be positioned in real-time using the mouse, but cannot be moved through any real obstacles.



(Figure 3.1 – Photograph of monitor showing a virtual cube passing under a real archway)

### 3.2.2 Light interactions

Virtual objects placed in a real environment should be expected to appear as if lit by the light sources that exist in reality and, ideally, all illumination interactions that occur between real objects should be evident between real and virtual entities.

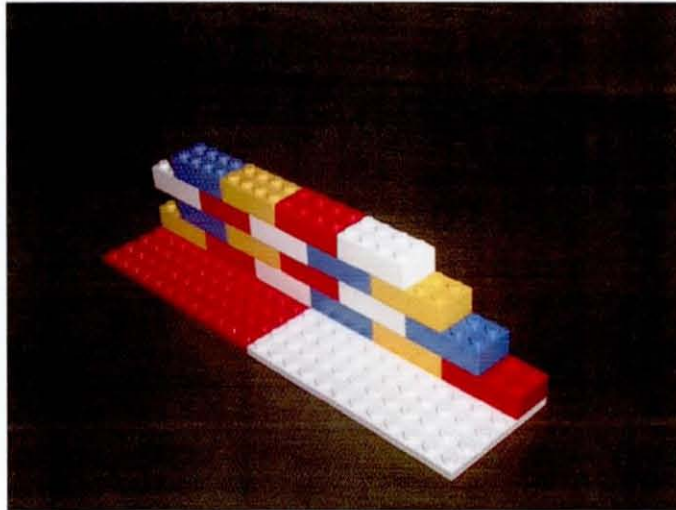
For full integration, as well as being able to compute the illumination of graphic objects by real entities, there is also a need to compute the illumination effect on the real scene of any graphic surfaces and light sources.

It is necessary to consider the consequences of interactions that may exist between:

- synthetic objects and synthetic objects (some of which may be light sources),
- synthetic objects and real objects (some of which may be light sources), and
- real objects and real objects (some of which may be light sources).

Some of these interactions pose significant problems. For example:

- there may be real objects that are not initially in view but that affect the illumination of the scene,
- virtual light sources may remove real shadows requiring reinstatement of real-world colour and texture that might not be known, and
- the indeterminable effects of hidden parts of real objects on virtual objects, such as the reflection of the back of a real object in a virtual mirror, or shadows caused by hidden parts of a real object illuminated by a virtual light source.



(Figure 3.2 - Real Lego™ wall with virtual brick inserted in top row; 2<sup>nd</sup> from the right)

There are also secondary illumination effects, such as reflections, shadows and transparency, which are influenced by both local and global illumination considerations. Glossy or mirrored surfaces in reality should reflect appropriately placed virtual objects, and the real should be reflected in the virtual. Real shadows may fall across virtual objects and virtual shadows across real objects. Reality should appear refracted through transparent graphics, and virtual objects refracted by the real. Atmospheric effects such as fog, smoke, heat haze, or just plain air should affect the appearance of virtual objects in the same way as they do real.





(Figure 3.3 - A virtual box reflected in a real mirror)

### *Specular reflections*

With respect to specular reflections, we would expect virtual objects to be reflected in real surfaces that are glossy or mirror-like. Likewise, we would expect to see real reflections in virtual surfaces. These reflections could range from the specular highlights caused by light sources to the mirrored reflection of object form and colour.

In the case of virtual reflections in real surfaces, there is necessity for the reflectivity of surfaces in the real scene to have been identified. For mirror-like surfaces there seems little problem in superimposing a virtual reflection. However, the situation becomes a little trickier when considering reflections in textured surfaces, where the virtual reflection would be expected to be perturbed by the real texture.

Producing appropriate reflections of real entities in virtual surfaces is a considerably more difficult problem. It would seem to require the mapping of a portion of the real scene onto the required virtual surfaces. However, the problem appears to become intractable when one considers that the reflected portion of the real world may include back facing or otherwise obscured surfaces for which appearance information cannot be gleaned from the current view.

### *Diffuse reflections*

The significant factor to consider, in relation to global illumination, is the effect that the mutual interactions of real and virtual objects would be expected to have on the balance of diffuse light radiation within an environment. The global illumination will be dependent on all the radiated energy due potentially to the effect that every point on every surface has on every other point on every other surface in the environment. This global model accounts for the ambient light within an environment and gives rise to the colour bleeding effects that are apparent when the colour of one surface influences the colour seen in a nearby surface, such as the pinkish tinge imparted to a white surface placed near a red. In computer graphics, a radiosity illumination model can be used to achieve such effects but, in the real world, the introduction of virtual objects into a real scene will upset the balance in a way that may be difficult to simulate. We would expect to see real objects, where appropriate, colour bleeding onto virtual surfaces and vice-versa. This is possible if we have a complete enough model of the geometry and reflectivities of the real world entities to allow unified application of a radiosity-rendering model to the creation of virtual objects. However, the complementary virtual-real interaction is not as straightforward since it requires modification of the appearance of real objects.

### *Shadows*

A further consequence of introducing virtual objects into a real scene is that we would expect real shadows, where they exist, to fall across virtual objects. Again, if we possess a model of the real-world geometry including its light sources this presents little problem. However, if our knowledge of the world is limited to what lies within a restricted view volume, we have incomplete information with respect to the global illumination model and resultant shadow effects.

Similarly, our expectation would be that virtual shadows, where appropriate, lie across real objects. Again, this is tractable if we have a complete world model, but more difficult where world knowledge is restricted. An even less tractable situation arises when virtual light sources are expected to illuminate real objects. Virtual



overlay could be used to replicate real surface areas where they are to appear brighter, but it is not clear how it might be possible to determine what real shapes, colours and textures underlie dissipated shadows.

### *Transparency*

We would expect opaque real objects to occlude the virtual and virtual objects to occlude real. In a simulation of realistic transparency effects, we would also expect real objects to be refracted through virtual. This requires distortion of reality, which could be achieved if a virtual overlay is constructed to include the refracted view of the appropriate parts of reality. The converse situation, where virtual objects need to be refracted through real, presents less difficulty provided we have a suitable model of relevant portions of the real world.

### *3.2.3 Real-world modelling requirements*

Although, in principle, when compared with virtual reality, augmented reality appears to offer the advantage that real-world objects do not need to be modelled accurately, the extent of this clearly depends on the requirements of the application. At a basic level, for simple overlay of annotations, it may be sufficient to locate them by modelling the position of just a few significant points in the real world. However, for more convincing real-world integration, a fuller model of the real scene is required. Depending on need, this may entail;

- a quality of geometric modelling sufficient to allow real-virtual collisions and occlusions to be determined,
- a model of the real-world illumination to the level of quality necessary to produce the required interactional effects such as shading and shadows,
- a model of the material nature of real-world objects to the degree needed to determine the effects of 'physical' interactions between real and virtual entities.

Hence, the apparent advantage of augmented reality with respect to the need for real-world modelling may, in practice, be restricted mainly to the fact that, although a real-world model is required, unlike virtual reality, it does not usually need to be rendered.

In addition to the above technical issues, for successful augmented reality, an appropriate way must be found of physically presenting a composite image to the user.

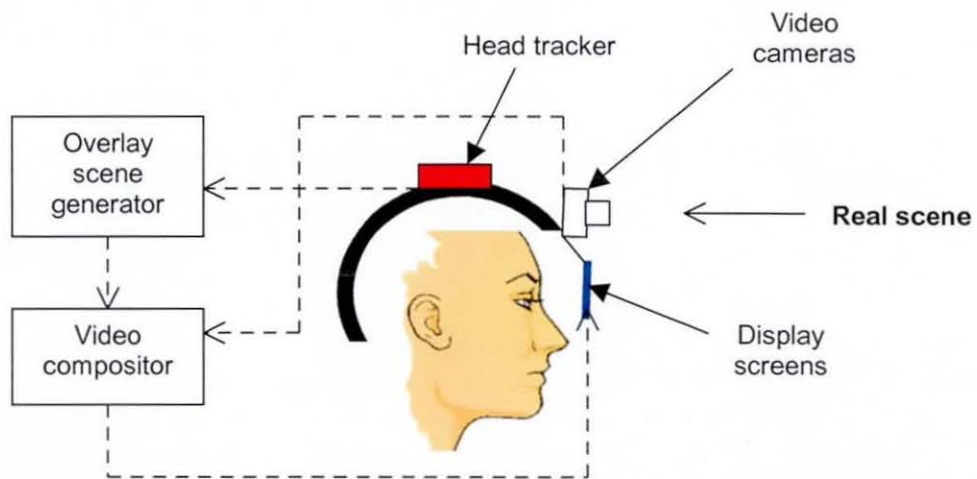
### 3.3 Displays

There are currently four approaches used for augmented reality display.

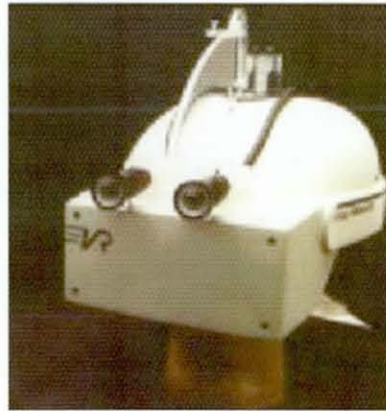
- Video see-through
- Optical see-through
- Monitor-based
- Projector-based

#### 3.3.1 Video-see-through display

A video-see-through display approach employs the same type of closed-view head-mounted display (HMD) as is used for virtual reality viewing. One or two head-mounted cameras provide the real-world view. The computer graphic elements are combined electronically with the camera images to form the composite. Figure 3.4 shows a conceptual diagram of a typical video-see-through augmented-reality system and figure 3.5, a photograph of an actual display.



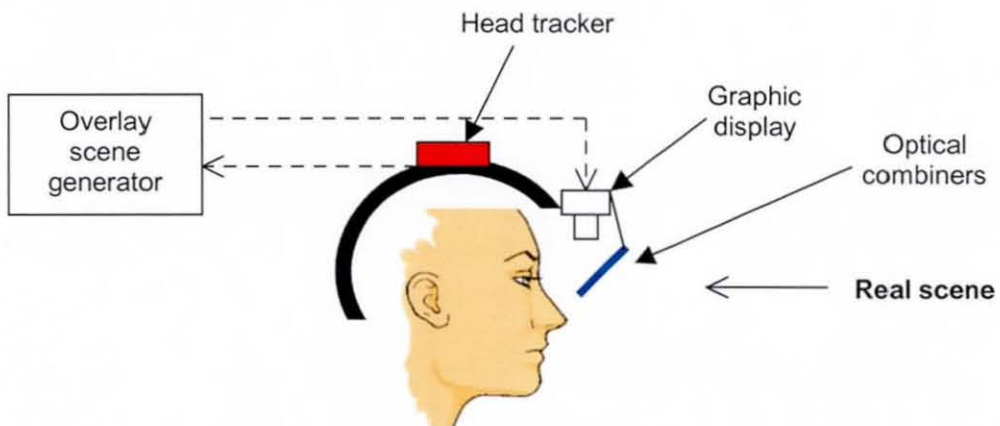
(Figure 3.4 – Conceptual diagram of video-see-through augmented reality system)



(Figure 3.5 - A video see-through HMD, UNC Chapel Hill)

### 3.3.2 Optical-see-through display

In contrast to video-see-through displays, optical-see-through systems place optical combiners such as part-silvered mirrors in front of the user's eyes. This enables the user to directly see the real world while, at the same time, computer-generated graphics can be overlaid on this view. The approach is similar to that used in the head-up displays (HUDs) used by military pilots. See the conceptual diagram in figure 3.6, and photograph in figure 3.7.



(Figure 3.6 – Conceptual diagram of optical-see-through augmented reality system)





(Figure 3.7 - An optical-see-through display made by Hughes Electronics)

Examples of this type of display include; Kaiser see-through version of Pro-View XL series [KAIS03] and Sony Glasstron [SONY03]. Microvision of Seattle [MICR03] has developed a Virtual Retinal Display™ in which the image is produced by scanning a low-power beam of coloured light onto the retina obviating the need for a display panel. With further development, this is anticipated to become small and lightweight enough to provide a comfortably wearable augmented reality display.

If superimposed graphic imagery is to be integrated successfully with real scenes, one of the fundamental issues to be considered is how to create superimposed virtual images that maintain a convincing impression of depth. Bajura et al. [BAJU92], in their experiments with the superimposition of ultrasound imagery on a real view of a human abdomen, identify this as a significant problem affecting the credibility of the graphic augmentation:

“Our experiment showed that simply overlaying synthetic images on real ones is not sufficient. To the user, the ultrasound images did not appear to be inside the subject, so much as pasted on top of her.” (p.208)

In an effort to address this problem, Bajura and his colleagues created a shaded polygonal pit to provide some simple depth cueing. However, the pit then occluded objects, such as the ultrasound transducer that should have appeared closer to the viewer, thus somewhat spoiling the illusion.

### 3.3.3 Monitor-based display

Monitor-based augmented reality merges computer graphics with real-time video streams. Commercially available systems such as that supplied by Princeton Video Image Inc. [PRIN03] are typically used in placing advertising logos into broadcast television transmissions; eg on a football pitch as illustrated in figure 3.8.



(Figure 3.8 – Monitor-based augmented reality from Princeton Video Image Inc.)

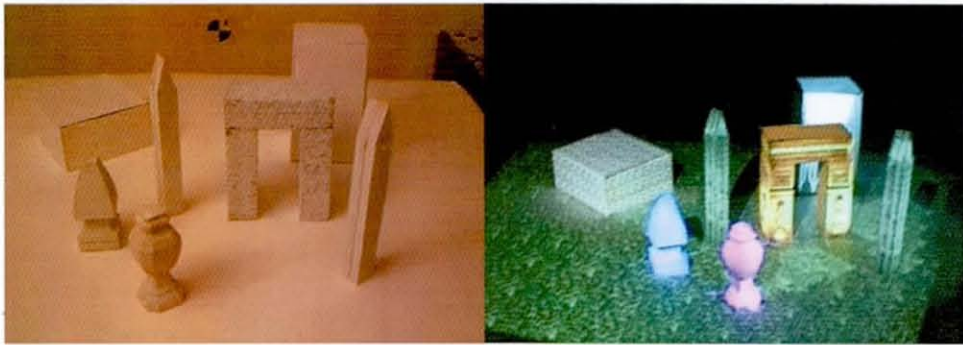
In this system, the cameras are equipped with sensors that measure the pan and tilt angles, and levels of zoom and focus. By comparing this information with a 3-D computer model of the stadium, a realistic video overlay is generated 30 times a second so that it appears to stay in place as the camera moves.

Overlaying the graphics onto the broadcast image uses a process similar to colour-keying, (discussed further in chapter 6) but instead of a single colour being used as the key, an operator needs to tell the system to overlay a range of field colours while excluding the colours of the players.

### 3.3.4 Projector-based display

Projector-based augmented reality uses image projection superimposed on physical objects. This has the advantage that, as the images are integrated directly into the

environment, many people can see the augmentation at the same time without need for specialised headgear. Animated projection can be via film or digital projector. The general approach is well established in the form of back projection for film compositing where, for example, an actor in the studio is made to appear as if in a different environment. Projection onto shaped objects is also sometimes used in museum exhibits. For example, at Madame Tussaud's in London, the author recalls seeing a waxwork figure with animated singing visage projected onto its face. Raskar et al. [RASK99] have also employed a similar principle in work that coordinates several ceiling-mounted projectors to texture and illuminate 3D objects. (See figure 3.9.)



(Figure 3.9 – Spatially augmented reality showing scene without and with projected imagery, UNC Chapel Hill)

### 3.3.5 *Optical versus video-see-through display*

As discussed above, head-mounted displays are either video-see-through or optical-see-through. There are comparative advantages and disadvantages of each, relating to:

- registration issues,
- intensity and resolution,
- colour range,
- safety,
- first-hand experience,
- compositing flexibility.



### ***Registration issues***

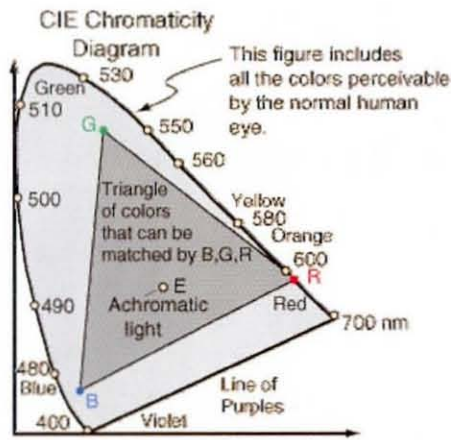
As illustrated by figure 3.5, in some video-see-through configurations, there is an offset between the user's eyes and the cameras. This introduces a displacement between what the cameras 'see' and what the user would normally see, which could create difficulties with respect to physical interaction with the real environment. No such displacement occurs in optical displays and, indeed, could be overcome in video-see-through systems using an optical arrangement similar to that shown in figure 1.5. With optical-see-through displays, the physical blending of the real and virtual worlds is achieved optically whereas, with video-see-through, digitising the video and compositing with the graphics introduces an additional time delay. However, with video-see-through displays, introducing a delay in the video stream can compensate modest temporal mismatches between the real and the virtual.

### ***Intensity and resolution***

The combiners used in optical-see-through displays inevitably reduce the intensity of light from the real world. On the other hand, with video-see-through, the resolution of the real world is limited to the resolution of the display panels which, at present, is far less than the resolution of the eye and hence a direct view of the world.

### ***Colour range***

With video-see-through systems the real scene is viewed via an electronic display causing some disadvantage in terms of the range of colours visible to the user. The problem being that colours which can be matched by combining a given set of RGB display primaries do not encompass all colours visible to the human eye. This can be shown on CIE Chromaticity diagram (see figure 3.10) in which three possible display device RGB values are shown plotted. The triangle joining these points encloses all of the colours that are within the gamut of these primaries; ie that can be produced by varying combination of them. No real three primary colours form a gamut that encompasses all visible colours.



(Figure 3.10 – CIE Chromaticity diagram showing RGB gamut, Georgia State University [GEOR03])

### **Safety**

In the case of video-see-through displays, a loss of electrical power will temporarily blind the user. With optical-see-through displays, power failure will only result in loss of the graphic overlay. This may be an important consideration in some applications such as surgical guidance.

### **First-hand experience**

For some envisaged applications of augmented reality the user will desire the first-hand experience of viewing the world as afforded by an optical see-through system. For example, visitors using an augmented reality display to view an archeological site apparently restored to its former glory are likely to want to view the world directly rather than wholly electronically processed.

### **Compositing flexibility**

A significant shortcoming of current optical-see-through displays is the lack of flexibility in compositing the real and the virtual. Virtual objects appear ghost-like and transparent such that a synthetic object nearer the viewer cannot be interposed properly in front of more distant real entities. Video-see-through techniques generally combine graphics and video via luminance or chrominance keying. Hence, unlike optical-see-through, there is the potential to modify the view of the original scene.



### **3.4 Discussion**

For some augmented reality applications, the overlaid graphic requirement is low, with only simple graphics required. However, the success of others may ultimately depend on the seamlessness with which synthetic graphics and reality can be merged. In light of the factors discussed above, for some applications there are in using an optical-see-through display rather than video-see-through. But, unfortunately, there are inherent problems with the current optical-see-through display arrangements.

Conventionally, in these displays, the computer-generated images are superimposed by reflection in a half-silvered mirror through which the real environment is viewed. This results in an overlay that always appears transparent, making it impossible to achieve convincing visual integration. For realistic compositing of virtual entities and a real environment, it is important that, when appropriate, virtual objects occlude or are occluded by real. Although some consideration has been directed towards achieving occlusion of virtual entities behind real objects [BREE95] [WLOK95] the fact remains that virtual objects cannot currently occlude real in optical-see-through systems.

For some applications, optical-see-through displays are to be preferred and the capability to modify the real-world view is an essential characteristic of any display system that is to allow virtual shadows to be cast over real surfaces or virtual light sources to appear to illuminate the real world. Unfortunately it is impossible to modify the appearance of any real world entities with current optical-see-through displays since they are viewed directly. Also, current see-through display technology is incapable of producing the visual occlusion and illumination effects needed for convincing real-world integration. This issue is dealt with in Chapters 6 and 7 where new optical-see-through display strategies are proposed and demonstrated. The perceptual issues relating to successful visual integration are discussed in Chapter 4.

## **4.0 Perceptual issues in augmented reality**

The technological issues to be resolved if useful augmented reality applications are to reach fruition are based, in large part, on the need to convince the human visual system that virtual objects are located correctly in real space; vertically, laterally and in depth. Carrying out real-world tasks will be impeded if the objects to be interacted with appear in the wrong position and at incorrect distances. The technological approaches to registration of projected views were discussed in Chapter 3. In this chapter the appropriate perception of depth is of main concern.

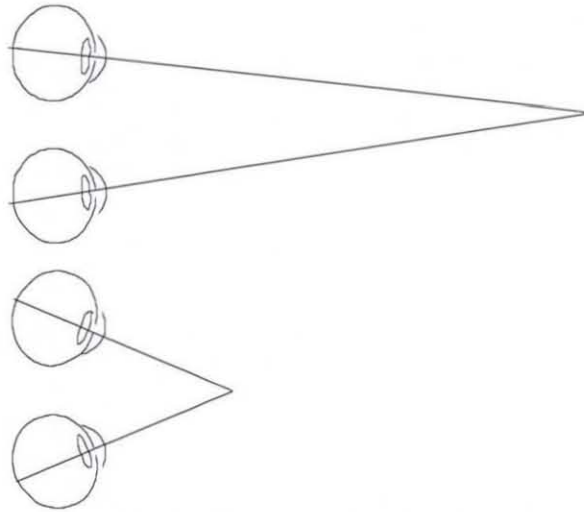
### **4.1 Depth perception overview**

There is an apparent paradox in our ability to perceive depth when our eyes have light sensitive surfaces that are two-dimensional. Depth perception cannot be explained by recourse to the simple eye-as-a-camera analogy that is often used to explain human vision, as what we actually perceive represents a more sophisticated construct than the small inverted retinal images initially received. We don't generally see things upside down, nor do we perceive double images despite the fact that we have two eyes, each receiving its own image. The British empiricist philosopher, Berkeley, in 1709, considered the problem of how we perceive depth [BRUC96] and his views remain largely dominant. Eye and brain work in concert to furnish our perceptions, and the perception of depth arises from this synergy. However, the empiricist view holds that the association of simple sensations received by the sense organs builds up all such complex ideas. Traditionally, there are a number of so-called cues that can account, at least in part, for the raw data required for depth perception. Some of these cues rely on binocular vision, some on relative motion, whilst others are able to provide information about depth to even a stationary, monocular viewer.

### 4.1.1 Convergence

Convergence is a binocular cue. When a distant object that is directly in front of the viewer is fixated, the eyes are positioned with the pupils central so that their respective lines of sight are almost parallel. However, for near objects the eyes rotate inward toward the nose so that the lines of sight converge toward the object. This convergence involves movement in the muscles controlling the position of the eyes and provides a potential cue to the relative depth of the fixated object.

When changing attention between two distant objects the degree of convergence will remain much the same so, not surprisingly, convergence has been found to be effective mainly for discriminating largish, near-far differences, not exceeding several metres [GRAH65].



(Figure 4.1 – Illustration to show convergence for a far and a near object)

### 4.1.2 Binocular disparity

Due to the spatial separation of the eyes, the retinal images differ, and this binocular disparity provides a further, and powerful, indication of depth. This was demonstrated convincingly by Wheatstone in 1838 [HABE80, p237] when he invented a mirror stereoscope with which he was able to prove that depth perception occurred



as a consequence of the disparity between the two views even in the absence of other depth information.

The power of binocular disparity alone is especially apparent in the recently popularised random dot stereograms first discovered by Bela Julesz at the Bell Telephone Laboratories in 1960 [ROCK84, pp61-62] in which three-dimensional shapes can be perceived within apparently random dot patterns.

The disparity between retinal images decreases in proportion to the square of the distance so, for distant objects, information about depth, gleaned from binocular disparity, must become increasingly less significant. Stereopsis is clearly not indispensable; closing one eye does not destroy the impression of depth in a scene.

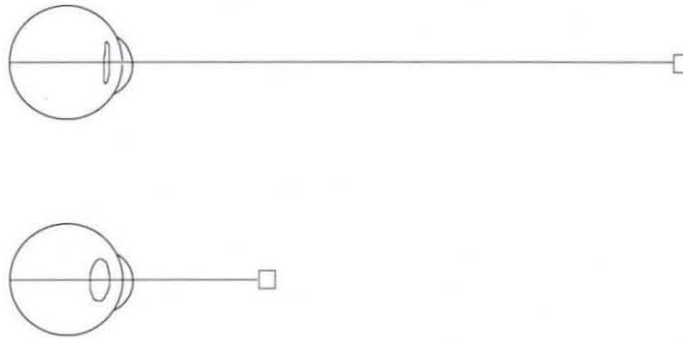
#### ***4.1.3 Motion cues***

Relative motions provide other cues to depth. When a viewer's head or eyes move laterally, the image of a nearby object passes across the retina faster than that of a more distant object. This motion parallax provides information about the relative depth of objects. As experiments with computer-generated two-dimensional random dot patterns, carried out by Rogers and Graham at Oxford University [ROGE79] have established, even motion alone can produce an impression of depth.

Motion has also been found to induce perception of depth through the kinetic depth effect such as can be produced in the shadow cast by a rotating object, and through the stereokinetic effect as perceived when eccentric circle patterns are rotated. Stereokinesis was first described in 1924 by Musatti [ROCK84].

#### 4.1.4 Accommodation

Convergence is not the only oculomotor depth cue available. The lens of the eye changes shape as objects at different distances are brought into focus.



(Figure 4.2 – Illustration showing lens shape during accommodation on a far and a near object)

However, accommodation appears to provide an effective depth cue at only relatively short distances. Objects that are not accommodated by the eye are blurred. Similarly, in photographs it might be possible to judge depth on the basis of focus. However, depth perception in photographs, drawings and paintings is more fully explained in terms of so-called secondary, or pictorial, cues; interposition, brightness, colour, shading, shadow and perspective.

#### 4.1.5 Interposition

If the contours of a surface appear to be partially obscured by another object, then the obscured object is usually perceived to be farther away. The partial covering of one object by another, or interposition, is a powerful depth cue.

Although it is clear that familiar shapes may be recognised as occluding one another, it is not so apparent how interposition of unfamiliar objects might be so easily determined. Surface texture and colour may support interpretation, and perception of form obviously has a role to play.



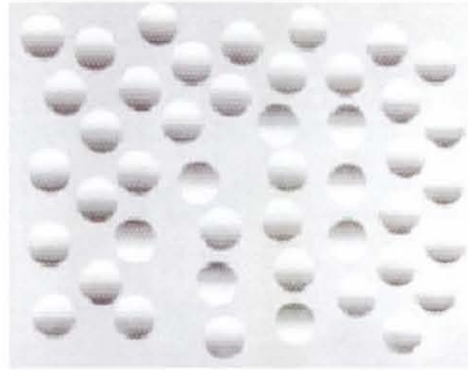
*(Figure 4.3 - A white rectangle; first unoccluded, then partially occluded by a chair. There is a strong impression of depth in the second image due to interposition)*

#### **4.1.6 Illumination and colour gradients**

Brightness, hue, colour saturation, shading and shadow are all capable of providing depth information. Generally, with increasing depth, objects appear less bright, with less distinct boundaries, and less saturated in colour. As the colour of an object becomes closer in quality to that of the background, the more it tends to recede into it. Attenuation of light reaching the eye from distant objects occurs due to the intervening atmosphere and, even in good visibility, distant objects appear tinged with blue due to impurities in the air. Attenuation with increasing distance will be more pronounced in foggy, smoky or dusty conditions.

The distribution of light and shade contains further potential depth cues. Shading gradients resulting from the illumination of surfaces give a three-dimensional appearance (figure 4.4), and attached and cast shadows provide cues to an object's position in depth. Similarly, reflections of objects in nearby surfaces can also provide cues to their relative depths.





(Figure 4.4 – Illustration showing depth cueing due to shading)

#### 4.1.7 Perspective cues

Linear perspective is a well-recognised cue to depth. Parallel edges in a scene appear to converge towards a vanishing point as they recede into the third dimension. In a related way, the relative apparent sizes of objects can be used to judge distance, even when object size is not actually known. As described by Emmert's Law [EMME81], objects of equal size at varying distances from the viewer project images with visual angles inversely proportional to their distance. This also ensures that less detail can be seen in distant objects.



(Figure 4.5 – Illustration showing depth cueing due to perspective.  
The rails appear to converge towards a distant vanishing point)

Additionally, all real surfaces have microstructure giving rise to some degree of texture, thus surfaces slanted away from the viewer project to produce a texture gradient in which the texture's granularity becomes finer as the distance from the viewer increases [GIBS50].



*(Figure 4.6 - Texture gradient across field of cereal)*

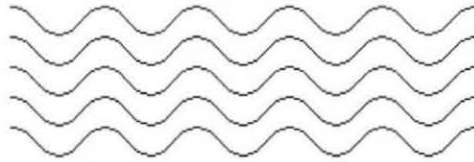
Due to linear perspective, the retinal stimulus of a given object relative to those of other objects in the visual field is a function of their distance relations. This has the effect that for an observer standing on flat ground and looking straight ahead, the ground extending towards the horizon appears higher in the retinal image as distance from the viewer increases. Thus farther objects, resting on the ground, appear to sit higher than those that are nearby.



*(Figure 4.7 - The distant chess piece appears higher in the image than one that is nearer the viewer)*



Probably, in a related way, contours can also provide a cue to depth. For example, the two-dimensional drawing in figure 4.8 gives some impression of depth.



(Figure 4.8 – Illustration showing depth cueing due to contours)

Depth cueing is also provided by the fact that atmospheric scattering effects cause very distant objects to appear less distinct than nearer objects; a phenomenon known as aerial perspective. See figure 4.9.



(Figure 4.9 – Illustration showing depth cueing due to aerial perspective)

#### **4.2 Depth perception in pictures and virtual environments**

Although a single cue can give rise to the perception of depth, in most realistic situations numerous cues are involved. Generally cues are mediated to provide the perception of a world that is stable with the true size, shape and relative positions of objects maintained regardless of the orientation or distance from the observer. When an observer moves closer to an object, the retinal image it subtends gets larger. However, this is perceived as a change in distance rather than a change in image size. This stable view of the world is generally referred to as visual constancy [GREG63].

In pictures and in situations where electronic displays are used, conflicting cues may be introduced, between which the visual system must attempt to arbitrate. In the case of a 2D picture, its image content may provide a scene with cues to 3D depth, whereas other cues may serve to inform the observer of its flatness of the image. 'Flatness cues' might include; binocular parallax cues, the presence of a visible frame, image surface texture, lack of colour depth, reflections at the image surface, as well as convergence and accommodation cues.

A number of authors have suggested strategies for pictorial viewing to improve the impression of depth. For example, Ames [AMES25] lists a number of ways to improve the impression of depth when viewing a picture, including;

- viewing with one eye,
- viewing at a distance,
- viewing through a small hole.

Schlosberg [SCHL41] has produced a similar list of ways in which flatness cues associated with 2D pictures may be reduced, including;

- monocular viewing,
- sufficient viewing distance to minimise accommodation effects,
- looking through a tube to block all light other than that coming directly from the picture.

Koenderink et al. [KOEN94] observe that such instructions as the above exist for improving pictorial depth, although point out that no objective verification of such claims exist. Hence they describe their own experiment in which pictorial relief is compared under monocular, 'synoptic' (ie with both eyes seeing the same image), and natural binocular viewing. They conclude that for observers with normal stereovision a painting is flat, but monocular and synoptic viewing reduces the 'flatness cues'.

In stereoscopic virtual environments, the convergence and accommodation appropriate for the visual display are in conflict. Thus, the stereoscopic stimuli may be set for a distance that differs from the one for which optimal focusing will occur. Roscoe [ROSC93] has demonstrated that pilots who are wearing see-through HUDs and viewing collimated virtual images will not focus their eyes on infinity, but instead toward their resting accommodation, which is approximately 1 metre distant. The result is that objects appear smaller than they should and hence interpreted as being further away.

Ellis and Bucher [ELL94] have shown that the judged position of a virtual object changes when it is superimposed on a real background using an optical-see-through display. They used a simple monochromatic wire-frame pyramid rotating about its vertical axis. An object with chessboard pattern was used as the real object. The apparent depth of the pyramid was measured using a LED pointer. Subjects were found to be able to match the LED position to the pyramid with reasonable consistency, although with larger than expected variability, when no background was present. In the presence of the real object at the same distance as the indicated depth of the pyramid, the pyramid was judged by most subjects to have jumped forward in space, so as to be in front of the chessboard. This seems logical since the chessboard did not occlude the pyramid. When the chessboard was moved forward a large amount, so that convergence clearly indicated that it was in front of the pyramid, many subjects reported that the chessboard appeared transparent.

### **4.3 Concluding remarks**

It is important to note that some depth cues can provide ordinal information only (ie sufficient to determine that one object is nearer than another), whereas other cues may have the potential for providing an absolute measure of depth. For example, occlusions provide ordinal information only, allowing the viewer to determine that one object is in front of another but not by how much. On the other hand, binocular disparity is capable of yielding absolute distance information for objects sufficiently near to the observer [LAND95]. Metric depth information can also be obtained



through motion parallax, as demonstrated by Ferris [FERR72] and Johansson [JOHA73], who have both shown that individuals are quite good at using this cue to judge absolute distances up to about 5 metres. Clearly, reliance on ordinal depth cues alone is not sufficient for an accurate absolute determination of depth. However, the effect of occlusion is striking in that, unlike other depth cues, its efficacy is not reduced over distance and its depth threshold, which is the minimum depth separation required for judging that two objects are at different distances, is lower than that of all other cues [CUTT97]. There is even some evidence [ANDE94] that stereopsis depends on partial occlusion.

In the real environment, occlusion and illumination effects provide forceful cues to depth. Unfortunately, it is these cues that are often not correctly catered for in augmented realities. The experiment described in Chapter 5 is designed to assess the relative efficacy of occlusion and shadow in the estimation of depth in an augmented scene.

## **5.0 The estimation of depth in augmented reality**

As discussed in chapter 4, interposition and shadows provide cues to depth in the real world but, as explained in chapter 3, these are cues that are impossible to provide in augmented reality based on current optical see-through display technology. Although problems caused due to lack of occlusion in these displays has been identified (see Section 3.3.2), the relative importance of shadows in improving depth estimation has not been established in this context.

This chapter describes an experiment designed to evaluate whether the presence of shadow improves estimation of depth more than is possible using interposition alone. The motivation here is to provide quantifiable justification for the effort entailed, not only in finding a solution to the optical-see-through display occlusion problem (discussed further in Chapters 6 and 7), but illumination issues also (which are the subject of Chapter 9).

### **5.1 Experiment aim**

Stereoscopic and dynamic depth cues play a significant role in the perception of depth but, in some video-see-through augmented reality systems, where such cues are not available, the impression of depth is not lost. Shutting one eye does not make our environment appear two-dimensional, and people who have lost an eye are able to interact with the three-dimensional world without too much difficulty; empirical evidence that testifies to the strength of monocular cues to depth. It is apparent that, if convincing integration of the real and virtual is to be achieved, it will be necessary to imbue graphic entities with a range of visual properties appropriate for their intended depth. Although non-illumination dependent cues such as interposition can provide a strong impression of depth, finer precision in estimation seems likely to depend on factors relating to illumination. Thus to establish a quantitative measure of the effectiveness of the depth cueing provided by an illumination dependent effect, an experiment was devised and conducted to assess the impression of depth



furnished by cast shadow, relative to the potentially strong depth cue of interposition in a composite real-virtual image.

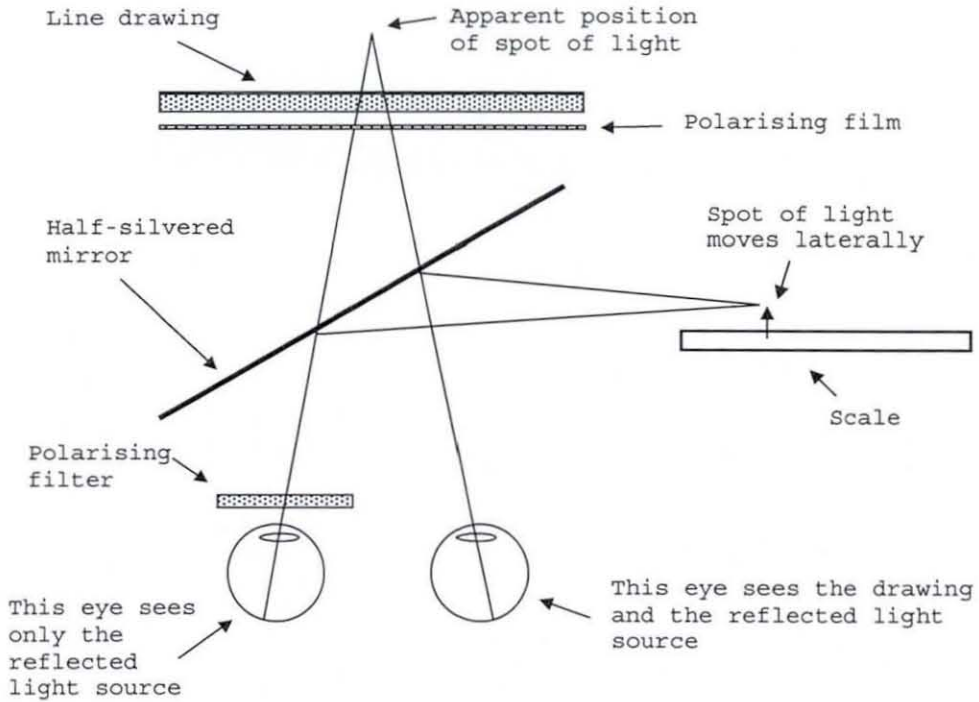
## **5.2 Experiment hypothesis**

The experimental hypothesis is directional, holding that the presence of cast shadow can provide a more accurate impression of depth in a static composite display than can be obtained from interposition alone.

To assess the efficacy of static cast shadows as a cue to depth, it is necessary to isolate these from other potential cues, and an experimental scheme must be devised that removes other cues from the virtual object allowing measurement of perceived depth where, in fact, no physical depth exists.

## **5.3 Previous virtual depth measurement experiments**

Measuring virtual depth, that is, measuring depth that is apparent to the perceiver but does not exist in reality, sounds by definition, an almost impossible task. However, an ingenious apparatus for this purpose was devised by Gregory [GREG77], and used to support his size constancy explanation of the Muller-Lyer illusion. A similar set-up (figure 5.1) was previously employed by Deregowski [DERE72] to measure cultural differences in the apparent depth of objects in line drawn pictures. As will be demonstrated, the apparatus also lends itself to the isolation of static, monocular depth cues arising from illumination effects.



(Figure 5.1 – Deregowski's apparatus for estimating apparent depth in pictures)

The original apparatus was based on the following principle. The illustration, whose apparent depth was to be measured, was covered with a sheet of polarising film. This was then viewed with one eye through a half-silvered mirror angled at  $45^\circ$  with respect to the illustration. Placed in front of the other eye was a polarising filter orientated at right angles to the first polarising sheet so that this eye could not see the illustration. A small light source was arranged orthogonally to the original line of sight so that both eyes could see its reflection in the partially silvered mirror in such a way that it appeared to emanate from within the illustration. Thus, the illustration was viewed monocularly whilst the reflection of the light source was viewed binocularly. As the light was moved backward and forward, its reflection appeared at varying depth within the illustration. In principle, the light could thus be adjusted using binocular vision until its apparent depth within the image coincided with the monocularly viewed object whose apparent depth, within the illustration, was to be determined.

#### **5.4 Comment on earlier experiments**

The original work carried out using the above apparatus has been criticised on two counts [HABE80]. First, on the grounds that not all flatness cues are removed effectively. Although viewing is monocular, other cues to flatness exist, allowing the perceiver to simply say that all parts of the image are equidistant. Second, that the pictures used were outline drawings which do not possess the range of luminance and spectral discontinuity that would provide information about the layout of space in a directly viewed scene or its photograph.

In addition, my own experience in replicating the apparatus suggests that, even using binocular vision, it is extremely difficult to judge the distance of a small spot of light moving backward and forward directly along the line of sight. Also, the presence of a polarising filter in front of one eye attenuates light to that eye only, potentially affecting any binocular judgement of depth.

#### **5.5 Apparatus modifications**

The criticism in relation to flatness cues is a difficult one to address fully. However, to avoid the cue afforded by a rectangular frame, the revised apparatus has been built using piping with a circular cross-section.

Accommodation provides another potential cue to flatness, as all parts of the screen are at the same focal distance, and a further possible complication is the tendency for objects to look smaller to people with normal binocular vision when viewed monocularly. The latter effect has been explained by Roscoe [ROSC84] as being caused when an occluded eye regresses toward resting focus, thus tending to move the seeing eye to a compromised point of accommodation. In the experimental apparatus, accommodation cues are, in effect, in some conflict as the real distances of computer screen and scale pointer are different. In defence of this potential source of criticism, there is evidence [RITT77] that convergence is dominant when in contest with accommodation, as was the case here, where convergence and binocular disparity were both cues available for positioning the pointer.



To reduce the influence of a further cue to flatness, i.e. screen texture, the computer display was cleaned thoroughly and, at the viewing distance used, it was felt that texture was not distracting or easily discernible. However, it must be noted that the display used had an inter-pixel-centre distance of 0.3mm, which, at the practical viewing distance of 500mm, subtends at the eye an angle of approximately 2 minutes of arc. This is twice the normal minimum angle assumed by the Snellen Chart eye test which sets the normal limit of acuity at 1 minute of arc, hence it must be expected that some screen texture may be noticeable to participants. There is some mitigation in that the screen intensity is attenuated slightly by the beam-splitter and filter, thus reducing acuity to some extent. Also, empirical evidence suggests that looking at a real scene through slightly textured glass does not seem to have any significant effect on depth perception and, in any case, any tendency to perceive the screen display as flat, by people using the apparatus, should be apparent in the measurements recorded.

The original criticism relating to the nature of the pictures used was avoided as the images for this experiment were based on a photograph of a real scene rather than simple line drawings. Indeed, the base image contained a range of pictorial depth cues, including; linear perspective, relative size, shading, shadows and texture gradient.

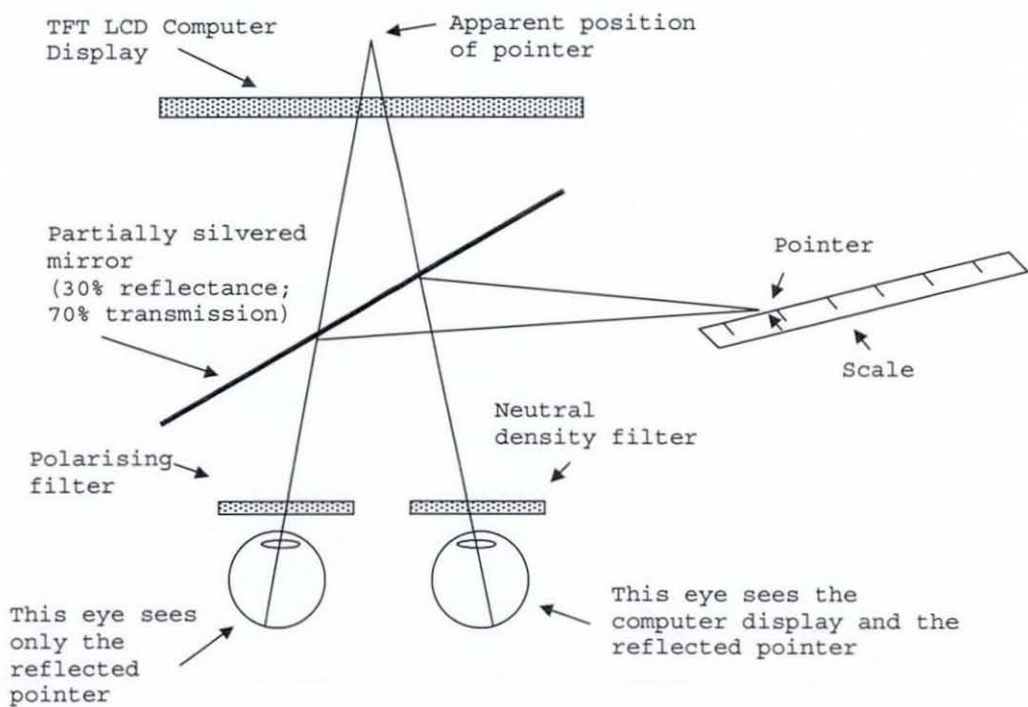
It was relative to this real view that the apparent depth of virtual objects was to be measured. Thus, in this case, the depth-positioning task could be made easier by positioning the scale obliquely to lie parallel with a line of perspective in the image. This allowed the viewer to see along the length of the scale as if aligned within the scene to measure along the real objects, and to position a small pointer along this scale. Although this arrangement introduces a lateral component that prevents direct measurement of absolute depth, the depth relative to the real objects can be determined more easily, and the problem of judging depth directly along the line of sight is avoided. To avoid the possibility of participants simply ignoring any depth cues and aligning the pointer with the virtual object according to its relative position

across the width of the flat display, the test object was designed as a virtual cylinder arranged to lie horizontally across the scene. The cylinder ends are squared off to remove any associated perspective cue.

The image was presented on a LCD computer screen that, in common with all such displays, was polarised. To counter the unbalanced attenuation of light to one eye by the polarising filter, a second identical filter was placed in front of the other eye to act as a matched neutral density filter, being rotated so it was not cross-polarised with the display.

Finally, to reduce the attenuation of the scene caused by a half-silvered mirror, a beam splitter with 70% transmittance and 30% reflectance was used rather than the half-silvered mirror of the original apparatus.

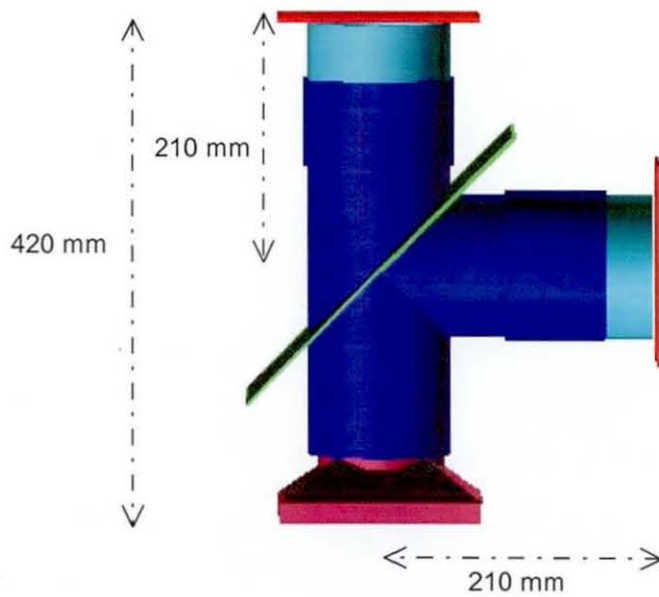
Fig 5.2 illustrates modified equipment set-up used.



(Figure 5.2 – Modified apparatus for estimating apparent depth in images)

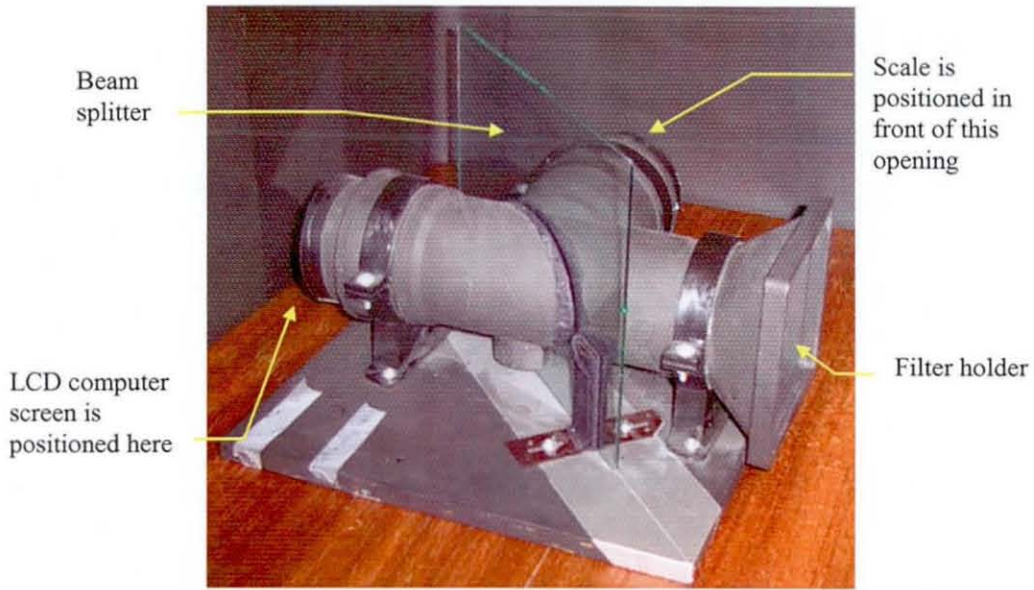


The viewing apparatus dimensions are shown in figure 5.3. These were selected to ensure that optical paths; reflected and transmitted are exactly the same. (The actual viewing apparatus is illustrated in figure 5.4. This equipment was later used as the basis of the prototype augmented reality display discussed in Chapter 7 of this thesis, so full details of its construction are provided there.)



(Figure 5.3 – Apparatus dimensions)

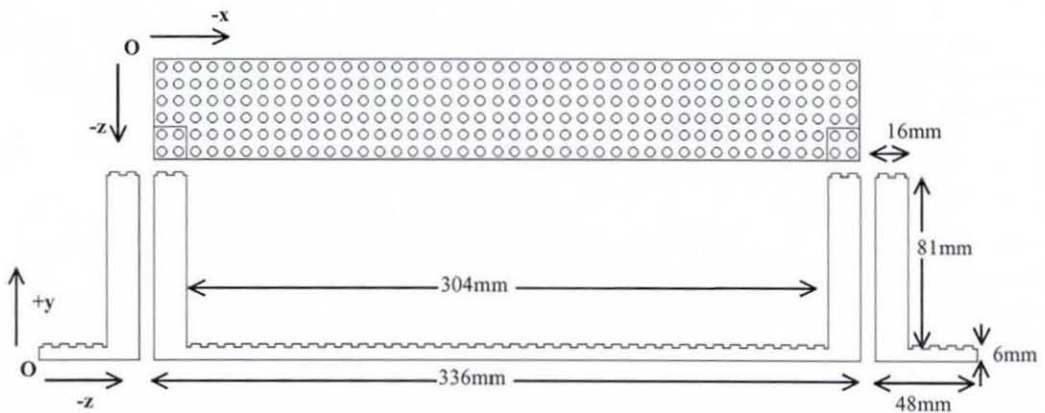
The mirror position was 210mm from the computer screen, and 210mm from the orthogonal pipe opening facing the measuring scale. The near end of the scale was pivoted at the pipe opening, 210mm from the mirror. Participants were required to sit such that the point halfway between their eyes was level with, and 290mm distant from the centre of the mirror. This distance was selected to be close enough to the apparatus to be able to see through it with both eyes, but far enough away for the participants not to knock it out of alignment accidentally.



(Figure 5.4 - Viewing pipe with beam splitter in place)

### 5.6 Test images

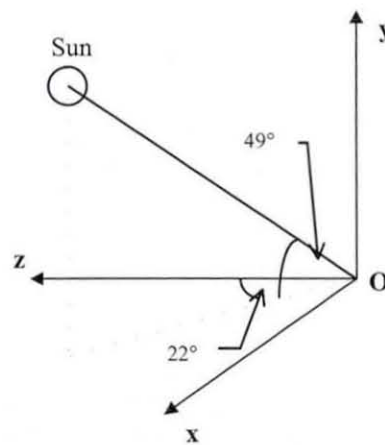
The experiment employed five test images with varying depth cues. The images were of a real scene, and the set comprised of a control image containing no computer graphics, along with four images that were superimposed with a computer graphic object with varying interposition and shadow cues. The real scene comprised a Lego™ base, with pillars at each end, illuminated by natural light. Figure 5.5 shows the object as a third angle projection. The world axes are shown using a right-handed world coordinate system with origin at **O**.



(Figure 5.5 – Real scene object design)

In world coordinates  $(x,y,z)$ , the camera was located at  $(470\text{mm},191\text{mm},-129\text{mm})$ , and was directed towards a 'look at' point  $(292\text{mm},6\text{mm},-36\text{mm})$ .

The light source was natural light from the Sun, which, relative to the scale of the object, can be assumed to have been at infinite distance. The angles relative to the world coordinate system are shown in figure 5.6.



(Figure 5.6 – Sun position)

The images were produced to provide a control image with no augmentation and four test images with each possible combination; interposition without shadow, interposition with shadow, no interposition but with shadow, and no interposition and no shadow. The images are reproduced as figure 5.7.

*Image0 (Control image):*

A control image with no superimposed graphic object.

*Image 1:*

A shaded graphic object superimposed without interposition of real entities in front of the virtual, with no cast shadow.

*Image 2:*

A shaded graphic object superimposed without interposition of real entities in front of the virtual, but with a superimposed cast shadow.

*Image 3:*

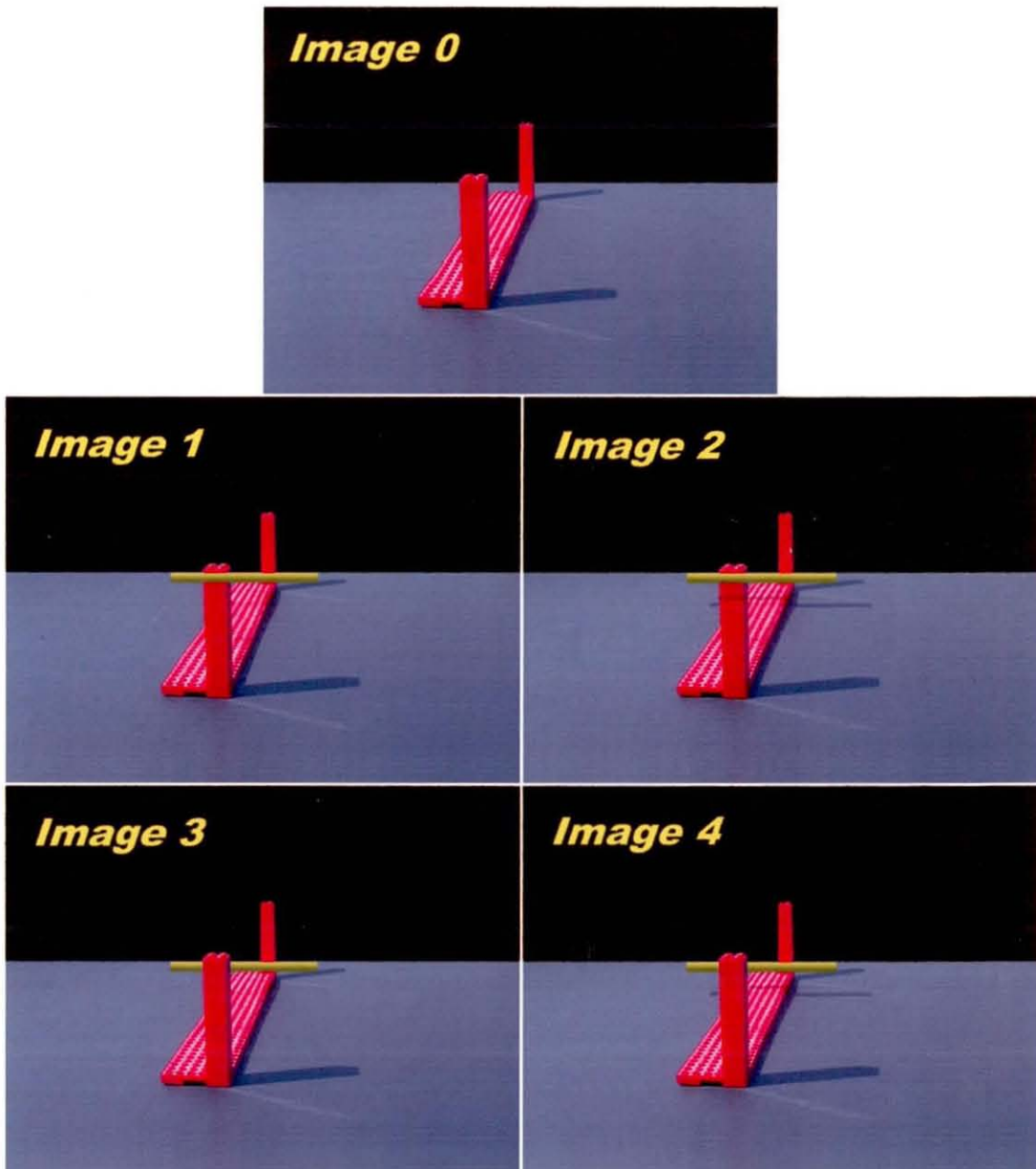
A shaded graphic object superimposed with interposition of real entities in front of the virtual, but with no cast shadow.

*Image 4:*

A shaded graphic object superimposed with interposition of real entities in front of the virtual, and with an interposed cast shadow.

|               | <b>Interposition</b> | <b>Shadow</b> |
|---------------|----------------------|---------------|
| <b>Image0</b> | n/a                  | n/a           |
| <b>Image1</b> | No                   | No            |
| <b>Image2</b> | No                   | Yes           |
| <b>Image3</b> | Yes                  | No            |
| <b>Image4</b> | Yes                  | Yes           |





(Figure 5.7 – Test scenes)

The original real-world image was obtained using a Canon PowerShot 600 Digital Camera set to a resolution of 640 x 480 pixels.

The virtual cylinder was created using ray tracing in a virtual world with distances proportional to those in the real world. The virtual camera view angle was set to 38° to match that of the digital camera. (See appendix A.) The 150mm long, 5mm diameter, cylinder was created with the axial centre of one end coincident with the



world origin, and the main axis running parallel to the world z axis. Its actual position with respect to the world was achieved by a translation of (-260mm, 25mm, 40mm). The ends were 'cut' square to remove cylinder-end perspective cues, and the cylinder was placed horizontally so that perspective in relation to the pillars would not influence depth estimation. Combining the real and virtual images using a digital editing package created the various composites.

For each of the test images, the participants were asked to judge the apparent depth of the virtual object relative to the near and far pillars. (See procedure description in section 5.10.)

### **5.7 Disparity adjustment**

An unexpected effect of the experimental arrangement was apparent if the view was alternated between the left eye and the right eye. Rather than the disparity shift of the measuring scale seeming greater at the end nearest the viewer, as would be expected, the greater movement appeared to occur at the far end of the scale.

It is believed that this occurred because the left eye alone sees the two-dimensional computer image, a left eye view of the inner walls of the piping, and a left eye view of the scale, while the right eye sees the two-dimensional computer image with no disparity as the screen is flat, a right eye view of the inner walls of the piping, and a right eye view of the scale with normal binocular disparity in relation to the left eye image. The computer image does not move in relation to the circular pipe end that is attached to the edge of its pipe opening. (The two pipe openings are visually superimposed as one is seen through the beam splitter and the other is reflected by it.) Because the computer image and the near end of the scale appear at fairly fixed positions in relation to the pipe openings, the difference between left and right eye images of the scale are seen as a far end lateral shift of the scale rather than the normally to be expected greater near end disparity. Figure 5.8 demonstrates the effect with superimposed lines to represent the left and right eye images of the scale.



(Figure 5.8 – Left and right eye-viewing effect on apparent scale alignment)

To prevent this creating a disparity between measurements taken with the left and right eyes, the scale was pivoted at its near end and two fixed positions were determined for the far end so that the scale could be moved between these locations to appear aligned properly with the real-world perspective for both the left and the right eye. The scale was fixed in one position for all left eye measurements, and the other for all right eye estimations.

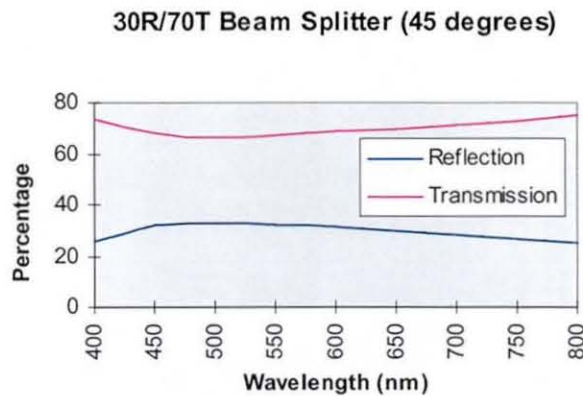
### 5.8 Participants

Twelve volunteer participants were selected on an opportunistic basis from the student population at Coventry University. There were 8 male and 4 female participants and all were aged between 19 and 24. To the best of their knowledge none had any known eye defects. No candidates with glasses were able to participate due to the polarising glasses that had to be worn, although those wearing contact lenses for normal vision correction were not excluded. For the sake of efficiency, and to counter individual characteristics that could confound experimental results within a set of image measurements, ‘within subjects’ design was used, with all participants experiencing all test conditions.

## 5.9 Additional precautions

### 5.9.1 Experimental equipment and environment

The part-silvered mirror was 254 x 356 x 3mm, soda lime glass, with 30% reflectivity and 70% transmission. In addition, the surface was MgF<sub>2</sub> coated to reduce surface reflections. The specified spectral performance was specified as being fairly constant across wavelengths as indicated by the graph in figure 5.9.



(Figure 5.9 – Spectral performance of beam splitter)

The closed-pipe system built for viewing had its inner surfaces painted matt black. A low intensity lamp was used to illuminate the scale pointer sufficiently for it to be lit evenly and seen clearly. The surrounding area was shielded with matt black paper and the experiment was carried out in a darkened room to prevent spurious reflections. Other precautions were built into the modified apparatus design as described in Section 5.5.

### 5.9.2 Visual anomalies

Individual differences between the visual systems of participants presented a potential source of error in this experiment. Therefore, precautions needed to be taken to ensure that all participants had good quality vision with no anomalies that could frustrate the analysis. There are many possible eye defects, both anatomical and pathological and, unfortunately, resources were not available for thorough eye testing of each participant. However, it was possible to identify beforehand the



specific visual faculties crucial to the experiment and to take steps to filter participants accordingly.

During the experiment, participants were expected to be able to judge, monocularly, relative depth in a flat image, at the same time as judging relative depth, stereoscopically, in the real world. Therefore, the following factors had to be considered in the selection of participants and in the experimental design.

1. Some reduction in visual acuity can be expected with increasing age. Physiological changes generally lead to slightly reduced visual acuity, altered colour vision so that blues are subdued and reds enhanced, delayed and reduced dark adaptation, and impaired recovery from dazzling glare [PARR89, p188]. To reduce such effects, only participants with no known uncorrected eye defects and aged between 19 and 24 were selected.
2. For any particular participant, differences between the left and right eye may cause the judgement to vary depending on which eye is used. People with anisometropia, have vision disturbed by a significant difference in refraction by each eye. As vision may still be very good in the other eye, the problem may have gone unnoticed. Another defect with potentially serious implications for this experiment is aniseikonia. This can give rise to a disturbance in spatial perception making the floor appear to slope or flat surfaces appear tilted. This can occur as a result of inherent differences in the visual system, such as a difference in the optical system or the length of each eye. As the differences may exist at birth or come on very gradually, in many cases, the visual system adapts to the difference, either tolerating it, or suppressing one eye [PICK89]. In both of these cases, it is possible for a prospective participant to be unaware of the problem, and thus not be able to admit to any visual defects. As a precaution, the experimental procedure was conducted twice for each participant with the cross-polarising filter over different eyes so that results were obtained for both left and right eye.

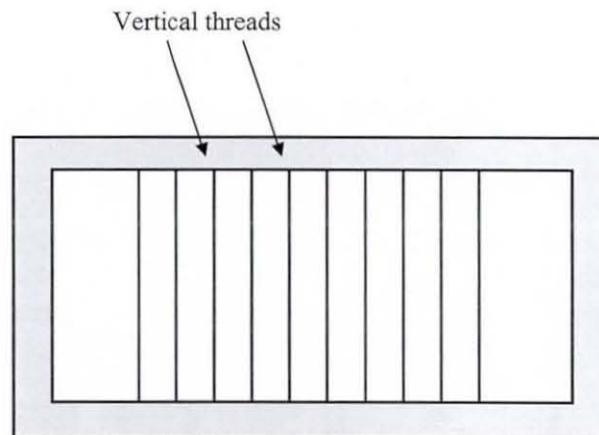
Significant differences between the two eyes could then be identified in the results and treated with due caution.

3. In order to make reasonable judgement of depth using, predominantly, the retinal disparity caused by binocular vision, participants needed to possess sufficient stereoscopic acuity. Stereoscopic acuity is actually a measure of an observer's ability to detect small differences in the distances of two objects, and is limited by the smallest amount of retinal disparity that can be perceived. The angular disparity threshold for stereopsis, also known as the stereo-acuity threshold, is in the range 2-10 arc seconds. Angular disparity  $\eta$  is given by;

$$\eta = P * \Delta D / D^2$$

where D is the distance of a fixated object from the eyes,  $\Delta D$  is the additional depth of another object beyond the fixated object, and P is the interpupillary distance.

A pre-test for stereoscopic acuity was carried out using a series of 10 vertical threads each set either forward or backward by  $\pm 4$  mm. (See figure 5.10.)



(Figure 5.10 – Stereo-acuity pre-test apparatus)



With their eyes 2 metres from, and orthogonal to, the vertical threads, participants were asked to state whether each was placed forward or backward. Assuming an interpupillary distance of 64mm, this gives:

$$\begin{aligned}\eta &= 64 * 8 / 2000^2 \\ &= 0.000128 \text{ radians} = 0.000128 * 180 * 3600 / \pi \text{ arc seconds} \\ &= 26.4 \text{ arc seconds}\end{aligned}$$

Ability to resolve about 26 arc seconds demonstrates a level of stereopsis above the expected threshold but double the acuity needed to resolve 1mm differences at the main experiment viewing distance of 500mm. Also with 10 threads there is only a 1 in 1024 chance (i.e. a probability of approximately 0.001) of obtaining the correct order by chance, and candidates who failed this test were not used in the main experiment.

### 5.9.3 Practice effects

Where each participant is to carry out measurements using each of the required test conditions, results can be affected by improving performance through practice or by the effects of fatigue. Such practice effects constitute a potential source of error associated with using 'within subjects' experiment design and need to be balanced as effectively as possible. In this experiment an attempt to achieve a satisfactory balance is accomplished by varying the selected order of test selection using Latin Squares, as described by Shaughnessy and Zechmeister [SHAU90]. Latin square sequencing ensures that:

- each test condition appears at every ordinal position equally often, and
- each test condition precedes and follows every other condition equally often.

To achieve this, Images 1 to 4 are assigned random numbers as follows:

|   |         |
|---|---------|
| 1 | Image 3 |
| 2 | Image 4 |
| 3 | Image 1 |
| 4 | Image 2 |

Using these assigned numbers, the 4x4 Latin Square becomes:

|             |             |             |             |
|-------------|-------------|-------------|-------------|
| 1 (Image 3) | 2 (Image 4) | 4 (Image 2) | 3 (Image 1) |
| 2 (Image 4) | 3 (Image 1) | 1 (Image 3) | 4 (Image 2) |
| 3 (Image 1) | 4 (Image 2) | 2 (Image 4) | 1 (Image 3) |
| 4 (Image 2) | 1 (Image 3) | 3 (Image 1) | 2 (Image 4) |

The 12 participants were assigned randomly to rows of this square with each row being used three times.

### 5.10 Procedure

#### *Step 1*

The experiment procedure was explained to the participant as described in appendix B.

#### *Step 2*

The experimental apparatus was set up with the light source switched off, the part-silvered mirror not yet in place.

#### *Step 3*

The participant was asked to sit in front of the screen and was helped to adjust the height and position of the chair so that his/her eyes were approximately level with, and 500mm distant from, the centre of the computer screen.

#### *Step 4*

The part-silvered mirror was positioned at the correct 45° angle, standing vertically, half-way between the participant and the computer screen.

*Step 5*

The participant was asked to look through the apparatus that was set so that one eye was cross-polarised with the screen, and to look straight at the centre of the computer display. The room lights were turned out.

*Step 6*

The control image, *Image 0*, was displayed on the computer screen and the pointer was positioned along the scale as close to the participant as possible before being illuminated.

*Step 7*

The participant was asked to direct the experimenter to move the pointer backwards and forwards, as necessary, until he/she was satisfied that it appeared to be at the same depth as the *near* pillar in the 'control' image; *Image 0*.

*Step 8*

The experimenter recorded the distance of the pointer along the scale.

*Step 9*

Steps 6 to 8 were repeated for the *far* pillar in the control image; *Image 0*.

*Step 10*

The experimenter repositioned the pointer to the scale point recorded for the *near* pillar, and the participant asked to confirm the accuracy of this repositioning. A test image was then displayed.

*Step 11*

The participant was then asked to direct the experimenter to move the pointer backwards and forwards, as necessary, until he/she was satisfied that it appeared at the same depth as the centre (near/far) of the virtual object.

*Step 12*

This measurement was recorded.

*Step 13*

Steps 10 to 12 were repeated for each test image in an order corresponding to the appropriate row of the Latin Square described above.

*Step 14*

Steps 5 to 13 were then repeated with the cross-polarising filter in front of the other eye, and the scale pivoted to the appropriate eye-perspective alignment position.

*Step 15*

Steps 1 to 14 were repeated for each of the participants, such that half the participants began with left eye cross-polarised, and half with the right eye .

**5.11 Results**

The results are shown in Tables 5.1 and 5.2. The scale used was 400mm long to extend slightly beyond the real length of the Lego™ object, and it was divided linearly into 40, 10mm units. Thus, in Table 5.1, all units are in centimetres.

| SubjectNo | Eye X-Pol | near pillar | far pillar | image 1 | image 2 | image 3 | image 4 |
|-----------|-----------|-------------|------------|---------|---------|---------|---------|
| 1         | left      | 11          | 32         | 10      | 10      | 28      | 17      |
| 1         | right     | 11          | 35         | 9       | 10      | 32      | 18      |
| 2         | left      | 13          | 26         | 13      | 12      | 15      | 21      |
| 2         | right     | 14          | 31         | 13      | 11      | 20      | 26      |
| 3         | left      | 10          | 27         | 11      | 10      | 11      | 19      |
| 3         | right     | 13          | 28         | 12      | 10      | 17      | 20      |
| 4         | left      | 11          | 33         | 9       | 8       | 18      | 22      |
| 4         | right     | 10          | 29         | 9       | 10      | 14      | 20      |
| 5         | left      | 9           | 25         | 8       | 8       | 17      | 21      |
| 5         | right     | 11          | 31         | 10      | 10      | 23      | 24      |
| 6         | left      | 12          | 30         | 11      | 11      | 16      | 26      |
| 6         | right     | 12          | 32         | 10      | 10      | 15      | 27      |
| 7         | left      | 13          | 35         | 11      | 12      | 21      | 26      |
| 7         | right     | 11          | 29         | 12      | 10      | 19      | 24      |
| 8         | left      | 12          | 28         | 10      | 8       | 18      | 19      |
| 8         | right     | 11          | 26         | 9       | 7       | 13      | 20      |
| 9         | left      | 10          | 31         | 8       | 9       | 18      | 26      |
| 9         | right     | 12          | 35         | 9       | 11      | 22      | 27      |
| 10        | left      | 14          | 34         | 12      | 12      | 26      | 22      |
| 10        | right     | 13          | 36         | 12      | 12      | 28      | 25      |
| 11        | left      | 12          | 29         | 11      | 9       | 18      | 26      |
| 11        | right     | 12          | 27         | 9       | 8       | 16      | 23      |
| 12        | left      | 11          | 26         | 10      | 9       | 19      | 17      |
| 12        | right     | 10          | 29         | 9       | 7       | 23      | 19      |

(Table 5.1)



Table 5.2 shows the data normalised with respect to the pillars so that the near pillar is at a distance of 0 units and the far pillar at 1 unit. Normalisation of each participant's results, for each eye, was achieved using the equation:

Normalised cylinder distance

$$= (\text{near pillar} - \text{apparent cylinder distance}) / (\text{far pillar} - \text{near pillar})$$

| SubjectNo | Eye X-Pol | image1 | image2 | image3 | image4 |
|-----------|-----------|--------|--------|--------|--------|
| 1         | left      | -0.048 | -0.048 | 0.810  | 0.286  |
| 1         | right     | -0.083 | -0.042 | 0.875  | 0.292  |
| 2         | left      | 0.000  | -0.077 | 0.154  | 0.615  |
| 2         | right     | -0.059 | -0.176 | 0.353  | 0.706  |
| 3         | left      | 0.059  | 0.000  | 0.059  | 0.529  |
| 3         | right     | -0.067 | -0.200 | 0.267  | 0.467  |
| 4         | left      | -0.091 | -0.136 | 0.318  | 0.500  |
| 4         | right     | -0.053 | 0.000  | 0.211  | 0.526  |
| 5         | left      | -0.063 | -0.063 | 0.500  | 0.750  |
| 5         | right     | -0.050 | -0.050 | 0.600  | 0.650  |
| 6         | left      | -0.056 | -0.056 | 0.222  | 0.778  |
| 6         | right     | -0.100 | -0.100 | 0.150  | 0.750  |
| 7         | left      | -0.091 | -0.045 | 0.364  | 0.591  |
| 7         | right     | 0.056  | -0.056 | 0.444  | 0.722  |
| 8         | left      | -0.125 | -0.250 | 0.375  | 0.438  |
| 8         | right     | -0.133 | -0.267 | 0.133  | 0.600  |
| 9         | left      | -0.095 | -0.048 | 0.381  | 0.762  |
| 9         | right     | -0.130 | -0.043 | 0.435  | 0.652  |
| 10        | left      | -0.100 | -0.100 | 0.600  | 0.400  |
| 10        | right     | -0.043 | -0.043 | 0.652  | 0.522  |
| 11        | left      | -0.059 | -0.176 | 0.353  | 0.824  |
| 11        | right     | -0.200 | -0.267 | 0.267  | 0.733  |
| 12        | left      | -0.067 | -0.133 | 0.533  | 0.400  |
| 12        | right     | -0.053 | -0.158 | 0.684  | 0.474  |

(Table 5.2)

### 5.12 Analysis

The experimental purpose was to check for significant differences in the estimation of depth between the four test images. All participants were tested for each condition using left and right eyes. Hence, analysis of results was carried out using Two-Way, Repeated Measures Analysis of Variance (ANOVA). The two factor design tests for

differences between the different levels of each treatment and for interactions between the treatments.

Factor A was set to be the Image with four levels (Image 1, Image 2, Image 3 and Image 4) and Factor B the Eye with two levels (Left and Right). The Dependent Variable was Distance.

The analysis was carried out using SigmaStat 3.0 software from SPSS Inc. [SPAA03].

### **Notes**

The results are calculated for each factor, and then between the factors.

### ***DF (Degrees of Freedom)***

The degrees of freedom are a measure of the numbers of subjects and treatments.

### ***SS (Sum of Squares)***

The sum of squares is a measure of variability associated with each element in the ANOVA table.

### ***MS (Mean Squares)***

The mean squares provide estimates of the population variances. The mean square for each factor is an estimate of the variance of the underlying population computed from the variability between levels of the factor.

### ***F Test Statistic***

The F test statistic is provided for comparisons within each factor and between the factors. If F is a large number, the variability among the means is larger than expected from random variability in the population, and you can conclude that the samples were drawn from different populations (i.e., the differences between the treatments are statistically significant).

### ***P value***

The P value is the probability of being wrong in concluding that there is a true difference between the treatments. The smaller the P value, the greater the probability that the samples are drawn from different populations. Traditionally, it can be concluded there are significant differences if  $P < 0.05$ .

In the analysis, the least square means and standard error of the means are displayed for each factor separately, and for each combination of factors.

Standard Error of the Mean provides a measure of the uncertainty in the mean.

### ***Multiple Comparison***

As a difference was found between the Image treatments, a multiple comparison table was computed. This was to determine exactly which treatments are different. In this case, the Holm-Sidak test was used.

### ***Holm-Sidak Test***

The Holm-Sidak Test can be used for both pairwise comparisons and comparisons versus a control group. It is more powerful than the Tukey and Bonferroni tests and, consequently, it is able to detect differences that these other tests do not.

When performing the test, the P values of all comparisons are computed and ordered from smallest to largest. Each P value is then compared to a critical level that depends upon the significance level of the test (set in the test options), the rank of the P value, and the total number of comparisons made. A P value less than the critical level indicates there is a significant difference between the corresponding two groups.

If the P value for the comparison is less than 0.05, the likelihood of erroneously concluding that there is a significant difference is less than 5%. If it is greater than 0.05, it cannot be confidently concluded that there is a difference.

*The results were as follows.*

**Two Way Repeated Measures ANOVA (Two Factor Repetition)**

**Data source:** Data 1 in Notebook 2

Balanced Design

Dependent Variable: Distance

| Source of Variation | DF | SS      | MS      | F       | P      |
|---------------------|----|---------|---------|---------|--------|
| Participant         | 11 | 0.403   | 0.0366  |         |        |
| Image               | 3  | 9.343   | 3.114   | 121.997 | <0.001 |
| Image x Participant | 33 | 0.842   | 0.0255  |         |        |
| Eye                 | 1  | 0.00141 | 0.00141 | 0.0996  | 0.758  |
| Eye x Participant   | 11 | 0.156   | 0.0142  |         |        |
| Image x Eye         | 3  | 0.0335  | 0.0112  | 2.355   | 0.090  |
| Residual            | 33 | 0.156   | 0.00474 |         |        |
| Total               | 95 | 10.935  | 0.115   |         |        |

The difference in the mean values among the different levels of Image is greater than would be expected by chance after allowing for effects of differences in Eye. There is a statistically significant difference ( $P = <0.001$ ). To isolate which group(s) differ from the others a multiple comparison procedure was used.

The difference in the mean values among the different levels of Eye is not great enough to exclude the possibility that the difference is just due to random sampling variability after allowing for the effects of differences in Image. There is not a statistically significant difference ( $P = 0.758$ ).

The effect of different levels of Image does not depend on what level of Eye is present. There is not a statistically significant interaction between Image and Eye. ( $P = 0.090$ )



Least square means for Image :

**Group Mean**

1.000 -0.0688

2.000 -0.106

3.000 0.459

4.000 0.598

Std Err of LS Mean = 0.0326

Least square means for Eye :

**Group Mean**

1.000 0.217

2.000 0.224

Std Err of LS Mean = 0.0172

Least square means for Image x Eye :

**Group Mean**

1.000 x 1.000 -0.0612

1.000 x 2.000 -0.0763

2.000 x 1.000 -0.0943

2.000 x 2.000 -0.117

3.000 x 1.000 0.461

3.000 x 2.000 0.458

4.000 x 1.000 0.562

4.000 x 2.000 0.633

Std Err of LS Mean = 0.0199

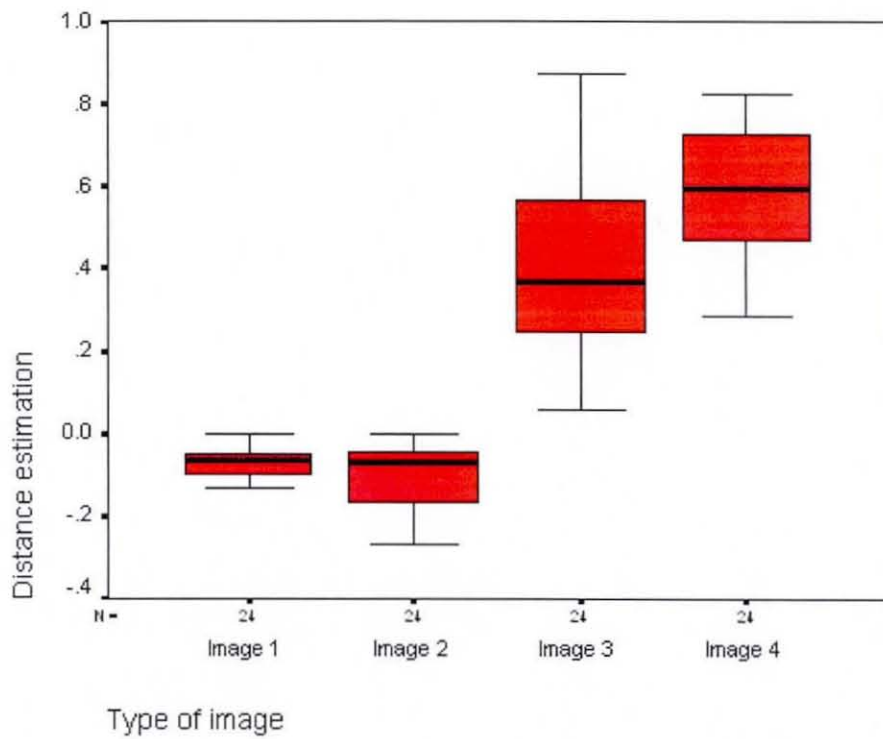
All Pairwise Multiple Comparison Procedures (Holm-Sidak method):

Overall significance level = 0.05

Comparisons for factor: **Image**

| Comparison      | Diff of Means | t      | Unadj P | Critical Level | Significant? |
|-----------------|---------------|--------|---------|----------------|--------------|
| 4.000 vs. 2.000 | 0.703         | 15.250 | <0.001  | 0.009          | Yes          |
| 4.000 vs. 1.000 | 0.667         | 14.452 | <0.001  | 0.010          | Yes          |
| 3.000 vs. 2.000 | 0.565         | 12.243 | <0.001  | 0.013          | Yes          |
| 3.000 vs. 1.000 | 0.528         | 11.445 | <0.001  | 0.017          | Yes          |
| 4.000 vs. 3.000 | 0.139         | 3.007  | 0.005   | 0.025          | Yes          |
| 1.000 vs. 2.000 | 0.0368        | 0.798  | 0.430   | 0.050          | No           |

Two Way Repeated Measures ANOVA Report Graph



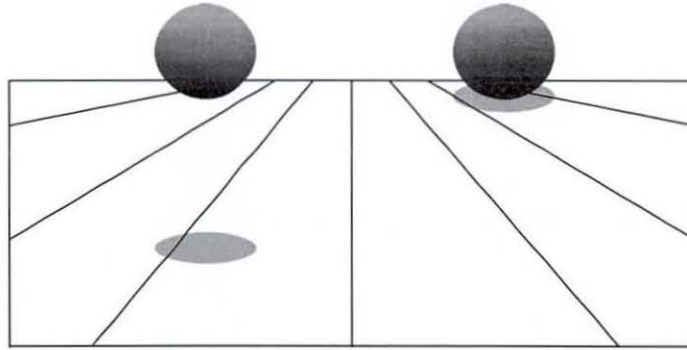
(Figure 5.11 – Box Plot showing estimated distances for each image)

### **5.13 Discussion**

The visual cues that enable humans to perceive depth are well documented. However, the use of these cues to convey depth information in computer-generated scenes is much less well researched. In virtual environments, certain depth cues may be missing altogether or cues may be in conflict. The experiment described in this thesis presented augmented images with some depth cues absent and others in conflict. Image 1 exhibited no occlusion of the virtual object by real entities and no cast shadow, Image 2 had occlusion and shadow conflicting with one another, Image 3 showed occlusion of the virtual object but no cast shadow, while Image 4 contained occlusion and shadow.

The results of the experiment show a significant difference in the mean values when comparing each test image, apart from between Images 1 and 2. The increased variance with respect to Image 2 when compared with Image 1 suggests that the inclusion of shadows without appropriate occlusion increased uncertainty of depth estimation. However, no participant estimated the cylinder in Image 2 to be at a greater depth than the first pillar, thus interposition clearly provided a stronger depth cue than the shadow. This is consistent with Johnston et al.'s [JOHN93] suggestion that in such situations, where there are conflicting cues, a vetoing mechanism applies so that the more dominant depth cue overrides the effect of the weaker cue.

Image 3 produced the greatest variance in depth estimation. This may be expected as in the absence of cast shadow there is ambiguity in relation to judging depth and height. To illustrate, in figure 5.12 [YONA78], most observers see the left object, with cast shadow, as being higher and closer to them than the right object, which has an attached shadow.



(Figure 5.12. Cast shadows influence perceived object elevation and depth)

In Image 3 there was neither cast nor attached shadow. In this situation depth estimation may be influenced by the cylinder height expected by the viewer. It is therefore possible that seeing images other than Image 3 first could have provided some prior expectation in relation to cylinder height. However, although Participant 1 (who did view Image 3 first) also gave the furthest distance estimate for Image 3, the results show no consistency in this effect.

Image 4, which incorporated interposition and shadow cues, was found to give an impression of depth with less variance than the corresponding image without shadow (Image 3) and gave a mean depth estimate that was closest to the 'real' cylinder normalised depth of  $252\text{mm}/320\text{mm} = 0.788$  although, at approximately 0.6, the estimated depth tended to be perceived somewhat closer to the viewer than the 'real' depth.

The experiment findings support the hypothesis that the presence of cast shadow can provide a more accurate impression of depth, in a static composite monocular display, than can be obtained from interposition alone. This provides some quantifiable justification for incorporation of pertinent illumination effects in augmented reality systems.

The following two chapters describe how these effects can be achieved in optical-see-through displays.



## **6.0 Compositing images in optical-see-through augmented reality**

Compositing multiple images into one has long been a requirement of the film industry where optical and now digital compositing are well established. However, compositing virtual elements into a real scene, viewed through an optical-see-through display, presents a new challenge. The real scene cannot be altered in the same way that a film or digital representation of it can. We cannot remove the parts we want to replace or alter the radiance or transparency of real-world objects. To address the problem, this chapter proposes a new optical-see-through augmented reality display concept and develops its underlying mathematics, leading to an algorithm for the automatic generation of display surface images.

### **6.1 The alpha channel**

Compositing one image onto another is a process much practised in the film industry. The classic method is to create the required spatial information using a travelling matte. This is a piece of film that is transparent where the overlaid object is to appear and opaque elsewhere. The complement of this is called a holdout matte, created by exposing another strip of monochrome film to the travelling matte. The holdout matte is placed in register against the background filmstrip while exposing to fresh colour film. This results in a film copy of the background unexposed in areas where overlay is to appear. These areas are then exposed to the overlay filmstrip through the travelling matte to produce the final composite.

Traditionally, there have been a number of ways devised for producing the original matte. One set of techniques [VLAH58] generates the matte filmstrip simultaneously with the overlay film. Another technique is to use chroma-keying. Typically, the overlay sequence is filmed against a bright blue background, and then a matte is generated that is transparent in areas corresponding to the blue background and opaque elsewhere [VLAH64]. Chroma-key compositing is relatively straightforward to carry out using a digital computer by replacing the chroma-coloured pixels in the overlay image with corresponding background-coloured pixels.



The above approaches to image compositing employ a simple binary decision at each image pixel; whether to use the overlay or the background colour. To provide greater flexibility, partial transparency is required. This allows mixtures of images where one can show through the other with varying degrees of transparency. Among other reasons, this is important at the edges between overlay and background to avoid aliasing.

The requirement for transparency led to the invention of the concept of the alpha channel by Smith and Catmull in 1977 [SMIT95]. This is now a fairly standard component of digital image storage formats, forming a fourth so-called alpha element in addition to the standard RGB (Red, Green and Blue) pixel colour description. The alpha channel is typically stored using the same number of bits as each colour channel so RGB images become RGBA (Red, Green, Blue and Alpha). The alpha channel provides a measure of the transparency of the pixel, with a value of 0 denoting full transparency.

This leads to an equation for linear interpolation between two images,  $A$  and  $B$ , where  $\alpha$  is the proportional influence of image  $A$  relative to  $B$ .

$$\alpha A + B - \alpha B$$

Porter and Duff [PORT84] noticed that efficiency would be improved if  $A$  is pre-multiplied by  $\alpha$  and stored as part of the image. As this product must be found for each of the RGB channels for each pixel in image  $A$ , a large number of multiplications is avoided at the time of compositing. The original motivation for this may have been one of efficiency at a time when multiplications were computationally expensive, however, forming images with pre-multiplied alpha is also conceptually closer to human visual understanding. If a pixel's alpha-value is 0, its pre-multiplied colours will be 0 and, in effect, the pixel will conceptually not exist. Using this approach, we can think of images as shaped rather than necessarily

defined by a rectangular boundary. This actually removes the notion of a travelling matte for compositing, as the shape of an object becomes integral to it.

As discussed in the following section, there is a fundamental problem in applying these principles to optical-see-through augmented reality systems.

## **6.2 Methods for applying an alpha channel to real scenes**

Compositing using traditional alpha-channel techniques is appropriate with respect to video-see-through augmented reality, where the virtual overlay is composited with a video representation of the real scene. In this situation, alpha-channel transparency can be applied to both the overlay and the video as required, hence compositing can be carried out. However, for optical-see-through systems there is no control at all over the transparency of the real scene and the effective range of alpha values for the overlay is restricted. This restriction arises because the overlay is viewed by reflection in a part-silvered mirror. A mirror providing a 50:50 reflection:transmission ratio, and assuming 8 bits are used for the channel, would allow an effective alpha range of only 0 to 127. Hence it is impossible for optical-see-through displays based on this principle to achieve a fully opaque overlay.

As part of this thesis it is proposed that this limitation be overcome by applying an alpha channel to the user's view of the real scene using physical masking. As will be shown, this can be accomplished by employing:

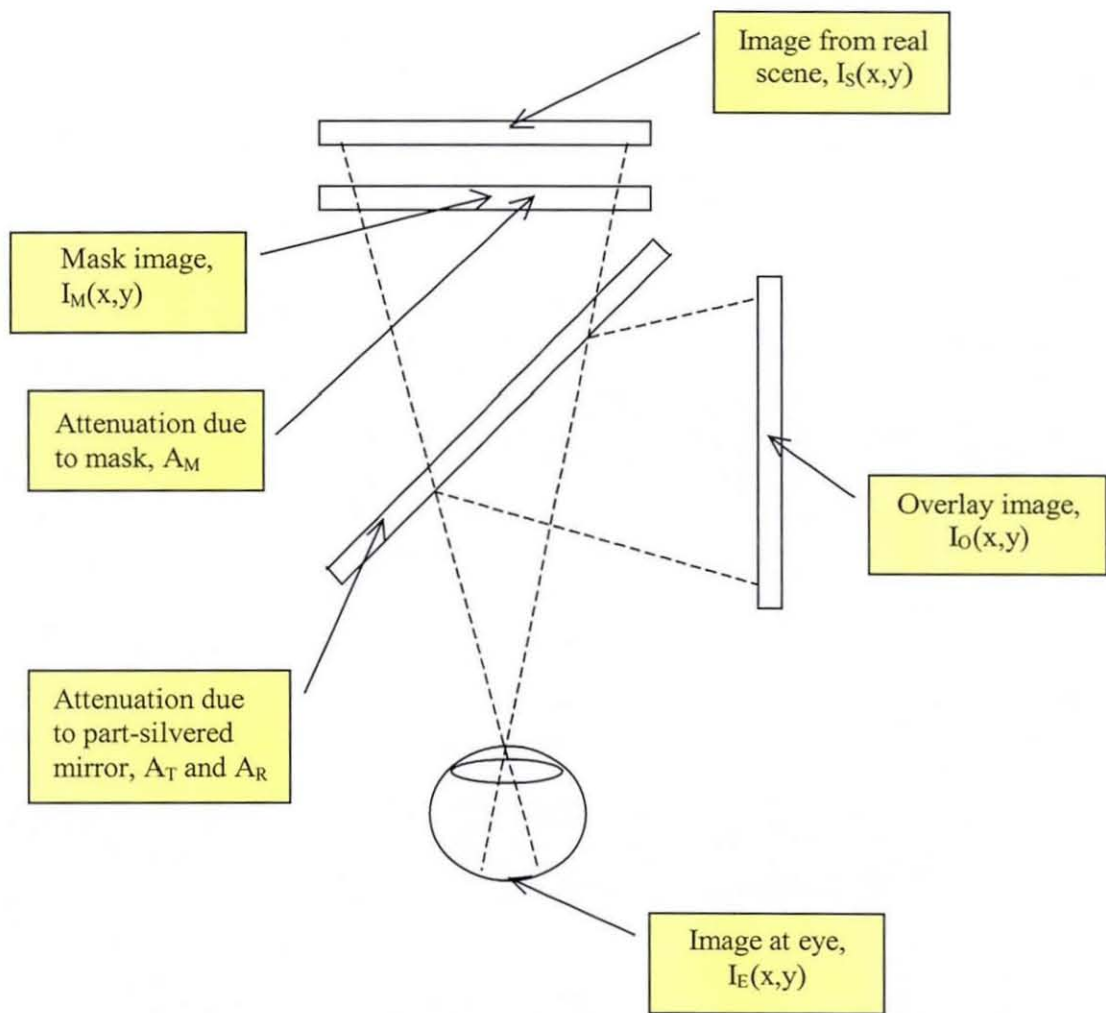
- an active filter panel through which the real scene is viewed,
- controlled, active illumination of the real scene, or
- selective reflection of real or virtual image elements.

For an optical-see-through augmented reality display it is initially proposed that a masking element be interposed between the real scene and the viewer as shown in figure 6.1. This arrangement could, for example, consist of a transparent LCD panel through which the real scene is viewed and can be actively masked as required. This

masked view is seen through a part-silvered mirror, which superimposes the reflection of the required virtual overlay. The mathematics for this is developed in the following section.

### 6.3 A mathematical model for see-through augmented reality

Figure 6.1 shows a possible see-through display arrangement with display surfaces identified.



(Figure 6.1 – Display surfaces for an optical-see-through system)



where:

$I_S(x,y)$  represents the intensity of light projected by the scene,

$I_M(x,y)$  represents the image intensity across the mask.

$I_O(x,y)$  represents the intensity across the computer-generated overlay image.

$I_E(x,y)$  represents the intensity of light projected into the viewer's eye.

$A_M$  represents the attenuation due to the masking panel substrate.

$A_T$  and  $A_R$  represent the attenuation due to the part-silvered mirror for the transmitted and reflected light respectively.

If we assume that all colours reflected from the original scene are constrained to those within the RGB gamut of the display system, (quite a big assumption but not unreasonable as a starting point) then the final intensity of each component projected towards the user's eye can be expressed by the following equation.

$$I_E(x,y) = [I_S(x,y) - ((2^{(\text{No. of bits per channel})} - 1) - I_M(x,y) \cdot A_M)]. A_T + I_O(x,y) \cdot A_R \quad \dots \text{eqn 6.1}$$

where

$$I_E(x,y) : 0 \leq I_E(x,y) \leq (2^{(\text{No. of bits per channel})} - 1) ;$$

In an ideal situation there would be no attenuation in the system components;

ie  $A_M = A_T = A_R = 1$ , giving an idealised form of equation 6.1:

$$I_E(x,y) = I_S(x,y) - [(2^{(\text{No. of bits per channel})} - 1) - I_M(x,y)] + I_O(x,y) \quad \dots \text{eqn 6.2}$$

(Note that, in practice, using a half-silvered mirror to reflect the overlay; ie  $A_T = A_R = 0.5$ , will cause a reduction of 50% in the overall intensity reaching the eye, however the relative intensities across the image will remain unchanged.)

Because the physical arrangement of the mask image and the overlay image constrain them to be capable of only subtracting and adding intensity respectively, they can be considered independently of one another. Thus for masking only, and assuming the ideal situation in which we can ignore display component attenuation;

$$I_E(x,y) = I_S(x,y) - [(2^{(\text{No. of bits per channel})} - 1) - I_M(x,y)] \quad \dots \text{eqn 6.3}$$

and for overlay only;

$$I_E(x,y) = I_S(x,y) + I_O(x,y) \quad \dots \text{eqn 6.4}$$

Rearranging equations 6.3 and 6.4 gives an equation for the mask:

$$I_M(x,y) = (2^{(\text{No. of bits per channel})} - 1) - (I_S(x,y) - I_E(x,y)) \quad \dots \text{eqn 6.5}$$

and for the overlay:

$$I_O(x,y) = I_E(x,y) - I_S(x,y) \quad \dots \text{eqn 6.6}$$



Utilising equations 6.5 and 6.6 provides the potential for automatic generation of appropriate mask and overlay components using the following algorithm. The algorithm assumes no attenuation in display surface substrates and, of course, that we know the appearance of the desired composite. In practice, the overlay image must also be mirrored horizontally to allow for reflection in the half-silvered mirror.

```

/* Algorithm to generate mask and overlay components,
   where;
   IS(x,y) = projected image from real scene
   IE(x,y) = projected image from composite scene
   IM(x,y) = mask image
   IO(x,y) = overlay image
*/

for each x
{
  for each y
  {
    difference1 = ( IS (x,y) - IE(x,y) )
    difference2 = ( IE (x,y) - IS(x,y) )
    if (difference1 < 0)
      IM(x,y) = 0
    else
      IM(x,y) = (2(bits per channel) - 1) - difference1
    )

    if (difference2 < 0)
      IO(x,y) = 0
    else
      IO(x,y) = difference2
  }
}

```

To illustrate, the following examples present calculation of mask and overlay components for individual pixels. An 8-bit representation for each of the RGB components is assumed.

Given that *pixelA* and *pixelB* are two pixels in the original scene, with RGB components as follows:

*pixelA*

$$R_{SA} = 161$$

$$G_{SA} = 165$$

$$B_{SA} = 161$$



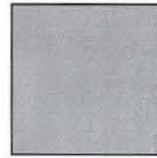
(Figure 6.2 – Original *pixelA*)

*pixelB*

$$R_{SB} = 160$$

$$G_{SB} = 161$$

$$B_{SB} = 163$$



(Figure 6.3 – Original *pixelB*)

Let's also assume that in the required augmented scene, *pixelA* is now in shadow while *pixelB* is to become brighter. For instance, *pixelA* may need to become:

$$R_{EA} = 49$$

$$G_{EA} = 49$$

$$B_{EA} = 47$$



(Figure 6.4 – *pixelA* in shadow)

and *pixelB*:

$$R_{EB} = 252$$

$$G_{EB} = 255$$

$$B_{EB} = 255$$



(Figure 6.5 – *pixelB* in highlight)

The darkening of *pixelA* is the responsibility of the corresponding pixel in the masking panel and the lightening of *pixelB* is accomplished through addition via the display overlay.

**Calculating the shadow mask**

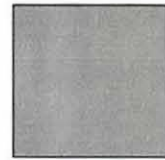
For calculating the shadow mask, equation 6.5 becomes:

$$I_M(x,y) = 255 - ( I_S(x,y) - I_E(x,y) ) \quad \dots \text{eqn 6.7}$$

where, if  $( I_S(x,y) - I_E(x,y) ) < 0$  then  $( I_S(x,y) - I_E(x,y) )$  is replaced by 0

Applying equation 6.7 to each of the RGB components, the corresponding shadow mask corresponding to *pixelA* becomes:

$$\begin{aligned} R_{MA} &= 255 - ( R_{SA} - R_{EA} ) = 255 - (161-49) = 255 - 112 = 143 \\ G_{MA} &= 255 - ( G_{SA} - G_{EA} ) = 255 - (165-49) = 255 - 116 = 139 \\ B_{MA} &= 255 - ( B_{SA} - B_{EA} ) = 255 - (166-47) = 255 - 119 = 136 \end{aligned}$$



(Figure 6.6 – *pixelA* shadow mask)

and the shadow mask corresponding to *pixelB*:

$$\begin{aligned} R_{MB} &= 255 - ( R_{SB} - R_{EB} ) = 255 - (160-252) \text{ resulting } 255-0 = 255 \\ G_{MB} &= 255 - ( G_{SB} - G_{EB} ) = 255 - (161-255) \text{ resulting } 255-0 = 255 \\ B_{MB} &= 255 - ( G_{SB} - B_{EB} ) = 255 - (163-255) \text{ resulting } 255-0 = 255 \end{aligned}$$



(Figure 6.7– *pixelB* shadow mask)

**Calculating the overlay**

For calculating the overlay using equation 6.6:

$$I_O(x,y) = ( I_E(x,y) - I_S(x,y) )$$

where, if  $( I_E(x,y) - I_S(x,y) ) < 0$  then  $( I_E(x,y) - I_S(x,y) )$  is replaced by 0

Applying equation 6.6 to each of the RGB components, the corresponding overlay pixel corresponding to *pixelA* becomes:

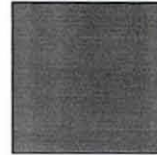
$$\begin{aligned} R_{OA} &= (R_{EA} - R_{SA}) = 49-161 \text{ resulting } 0 \\ G_{OA} &= (G_{EA} - G_{SA}) = 49-165 \text{ resulting } 0 \\ B_{OA} &= (B_{EA} - B_{SA}) = 47-166 \text{ resulting } 0 \end{aligned}$$



(Figure 6.8 – *pixelA* overlay)

and for *pixelB*:

$$\begin{aligned} R_{OB} &= (R_{EB} - R_{SB}) = 252-160 = 92 \\ G_{OB} &= (G_{EB} - G_{SB}) = 255-161 = 94 \\ B_{OB} &= (B_{EB} - B_{SB}) = 255-163 = 92 \end{aligned}$$



(Figure 6.9 – *pixelB* overlay)

Similarly, for a coloured pixel, *pixelC*, with the following RGB values.

*pixelC*

$$\begin{aligned} R_S &= 255 \\ G_S &= 102 \\ B_S &= 0 \end{aligned}$$



(Figure 6.10 – *pixelC* original)

If the required intensity in shadow is to be:

$$\begin{aligned} R_E &= 153 \\ G_E &= 51 \\ B_E &= 0 \end{aligned}$$



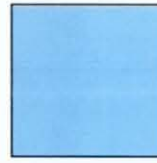
(Figure 6.11 – *pixelC* in shadow)

calculation of the shadow mask gives:

$$R_M = 255 - (255 - 153) = 255 - 102 = 153$$

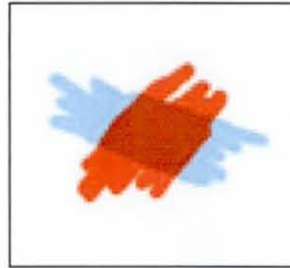
$$G_M = 255 - (102 - 51) = 255 - 51 = 204$$

$$B_M = 255 - (0 - 0) = 255$$



(Figure 6.12 – **pixelC** shadow mask)

The result of filtering the colour of **pixelC** with this colour mask is shown in the swatch shown in figure 6.13:



(Figure 6.13 – **pixelC** shadow masking colour swatch)

If the required intensity were to be brighter, for example:

$$R_E = 255$$

$$G_E = 204$$

$$B_E = 0$$



(Figure 6.14 – **pixelC** in highlight)

the calculated overlay pixel becomes:

$$R_O = (255 - 255) = 0$$

$$G_O = (204 - 102) = 102$$

$$B_O = (0 - 0) = 0$$



(Figure 6.15 – **pixelC** overlay)

The result of overlaying the colour of **pixelC** with this colour is shown in the next swatch:





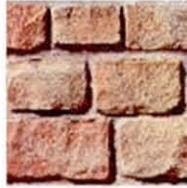
(Figure 6.16 – **pixelC** highlight overlay colour swatch)

#### 6.4 Error effects

In theory, where a virtual object is to occlude a real, the above algorithm should produce satisfactory results. In effect, adding or subtracting intensity from the background creates the colour of the overlaid object. Therefore, as a virtual object moves in front of a real background, the overlay to produce it constantly needs to change in a chameleon-like fashion, modifying the background colours to make the desired colour of the overlying virtual object. However, in practice, imperfect RGB approximation of real-world colour and attenuation at display surfaces prevents perfect reproduction of the overlaid object colour. Thus, for example, what might be intended to be a flat colour virtual rectangle in front of real patterned background is likely to show through the effects of that background.

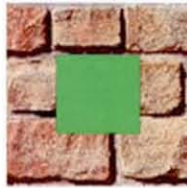
The following example shows a real brick wall as background and a virtual green rectangle is to be superimposed.

The real background:



(Figure 6.17 – Real background)

and the desired composite:



(Figure 6.18 – Desired composite)

The calculated mask for this augmentation is:



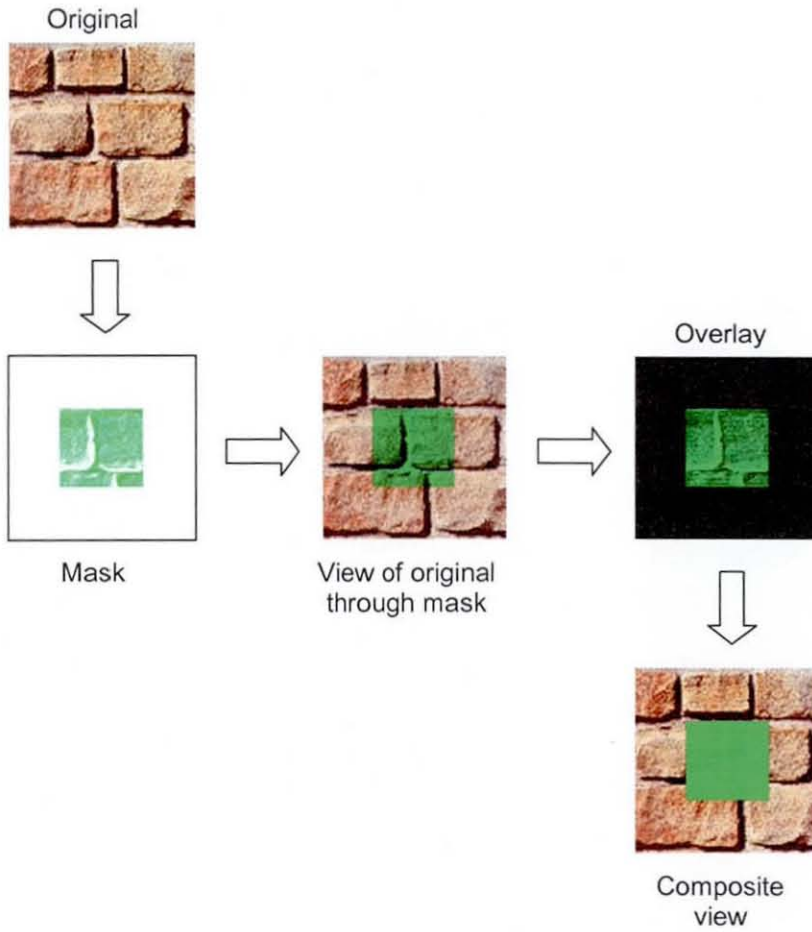
(Figure 6.19 – Calculated mask)

and the overlay:



(Figure 6.20 – Calculated overlay)

The image pipeline is illustrated in figure 6.21.



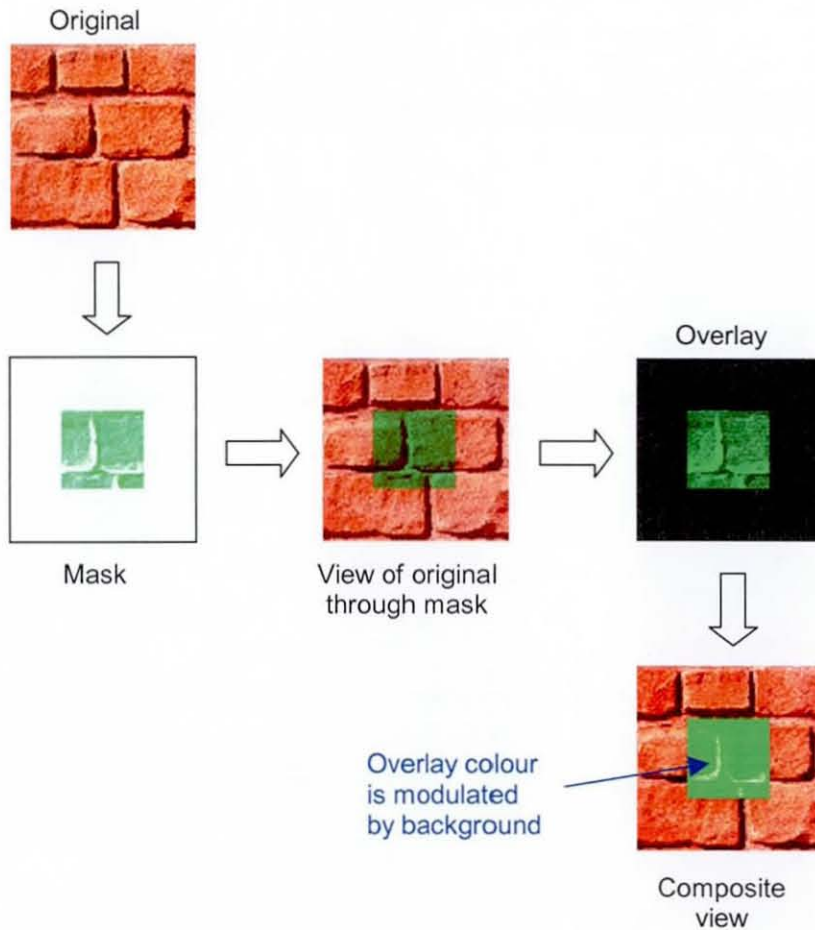
(Figure 6.21 – Image pipeline showing ideal compositing)

However, if we assume that the intensity of the real wall is given a slight red shift giving an actual original background appearance:



(Figure 6.22 – Real background with red shift)

The pipeline becomes as shown in figure 6.23:



(Figure 6.23 – Image pipeline showing error in composite due to colour inaccuracy)

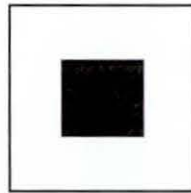
The net effect is errors, which allow the background to modulate the overlay giving an appearance reminiscent of transparency.

Clearly such problems have a much more disruptive visual effect in areas where virtual objects must occlude real. In other areas where real background is to be modified by addition of virtual shadow, colour bleeding or illumination, the acceptable visual tolerance of error is likely to be greater. It is therefore necessary to add to the mask, a solid black area of occlusion in areas corresponding to the projected silhouette of opaque virtual objects. In effect, this leads to the need to

create to a holdout matte; similar to that used in classic film-industry compositing as discussed at the beginning of this chapter.

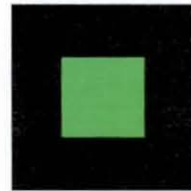
In our rectangle-over-brick example, the holdout matte takes the form of a solid black rectangle used as a mask to fully occlude the background beneath the overlay object.

In this example, the mask becomes:



(Figure 6.24 – Holdout mask)

and the overlay image:



(Figure 6.25 – Overlay for use with holdout matte)

This requires a modification of the earlier algorithm to incorporate a relative depth check between each projected pixel of the virtual overlay and its corresponding real background pixel. If the corresponding point on a virtual object is found to be closer to the viewer's eye than the background, the corresponding mask pixel must be set to black. Thus the holdout matte is a binary image (ie comprised of pixels that are 0 or 1) representing the shape of occluding areas of overlay objects.

Non-occluding effects that need to be produced by the mask can be calculated as before, with the final mask image being the product of multiplying  $I_M(x,y)$  and  $I_{\text{MATTE}}(x,y)$ .

A modification must also be made to the overlay image calculation, so that:



$$I_o(x,y) = I_E(x,y) - (I_S(x,y) * I_{MATTE}(x,y)) \quad \dots \text{eqn 6.8}$$

Modifying the algorithm accordingly;

```
/* DG(x,y) represents the depth buffer value for scene augmentation
projection at pixel coordinates (x,y).
DS(x,y) represents the depth buffer value for original scene
projection at pixel coordinates (x,y).
*/
```

```
for each x
{
  for each y
  {
    if DG(x,y) < DS(x,y) //ie augmentation is to
                          occlude reality because
                          virtual element is closer to the
                          viewer
      IMATTE(x,y) = 0 //solid black occlusion at
                      this pixel
    else
      IMATTE(x,y) = 1

    difference1 = ( IS (x,y) - IE(x,y) )
    difference2 = ( IE (x,y) - IS(x,y)* IMATTE(x,y) )

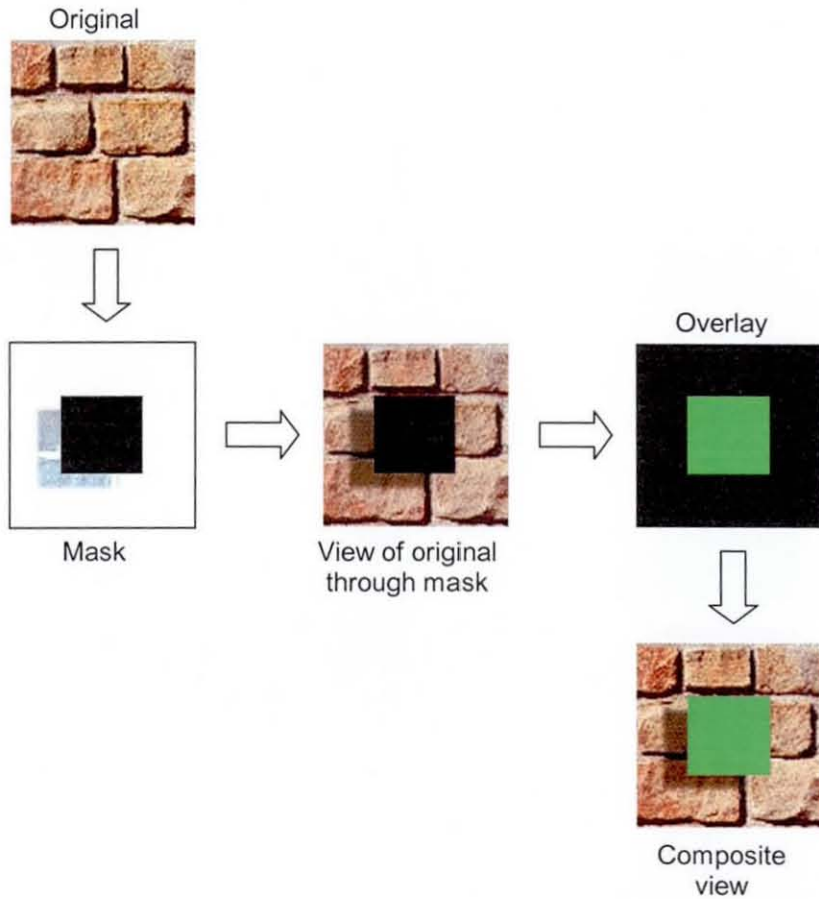
    if (difference1 < 0)
      IM(x,y) = (2(bits per channel) - 1) - 0
    else
      IM(x,y) = (2(bits per channel) - 1) -
                  difference1 )

    IM(x,y) = IM(x,y) * IMATTE(x,y)

    if (difference2 < 0)
      IO(x,y) = 0
    else
      IO(x,y) = difference2
  }
}
```

In the following example, the green rectangle is to be set in 3D space in front of the real background so that it casts its shadow on the bricks. The mask incorporates

shadow and occlusion matte (ie  $I_M(x,y) * I_{MATTE}(x,y)$ ). Figure 6.26 shows the image pipeline.



(Figure 6.26 – Image pipeline incorporating holdout matte)

## 6.5 Conclusion

In this chapter a mathematical model to describe compositing images in optical-see-through augmented reality displays has been developed. The practical realisation of

this model depends, in part, on a new image surface being added to existing display arrangements. The purpose of this image surface is to effectively add an alpha channel to the user's view of the real environment. In practice, attenuation will arise at the display surfaces and constraint on the gamut of colours in the original scene cannot be realistically assumed. This means that colour errors will occur in the composite. However, the most serious consequence, that is the production of background patterning in the appearance of virtual opaque objects, can be avoided through the expedient of adding a black holdout matte to the mask with shape corresponding to that of the virtual object.

The following chapter describes the design, construction and demonstration of display arrangements incorporating a real-world alpha channel.

## **7.0 Alpha-channel masking in optical-see-through displays**

As discussed in Chapter 3, current optical-see-through augmented reality displays produce a graphic overlay that is inherently transparent. It is, therefore, not possible to display virtual objects that occlude a real background. Neither can a real view be modified in the ways that would be needed if it is to show virtual shadows across real surfaces, or to apply virtual light sources to illuminate real objects. To address these limitations, the concept of applying an alpha channel to views of the real world has been developed in chapter 6 while, in this chapter, possible approaches to hardware implementation are proposed.

Real-world alpha-channel masking can be implemented in three ways. These are by:

- transmission
- projection
- reflection.

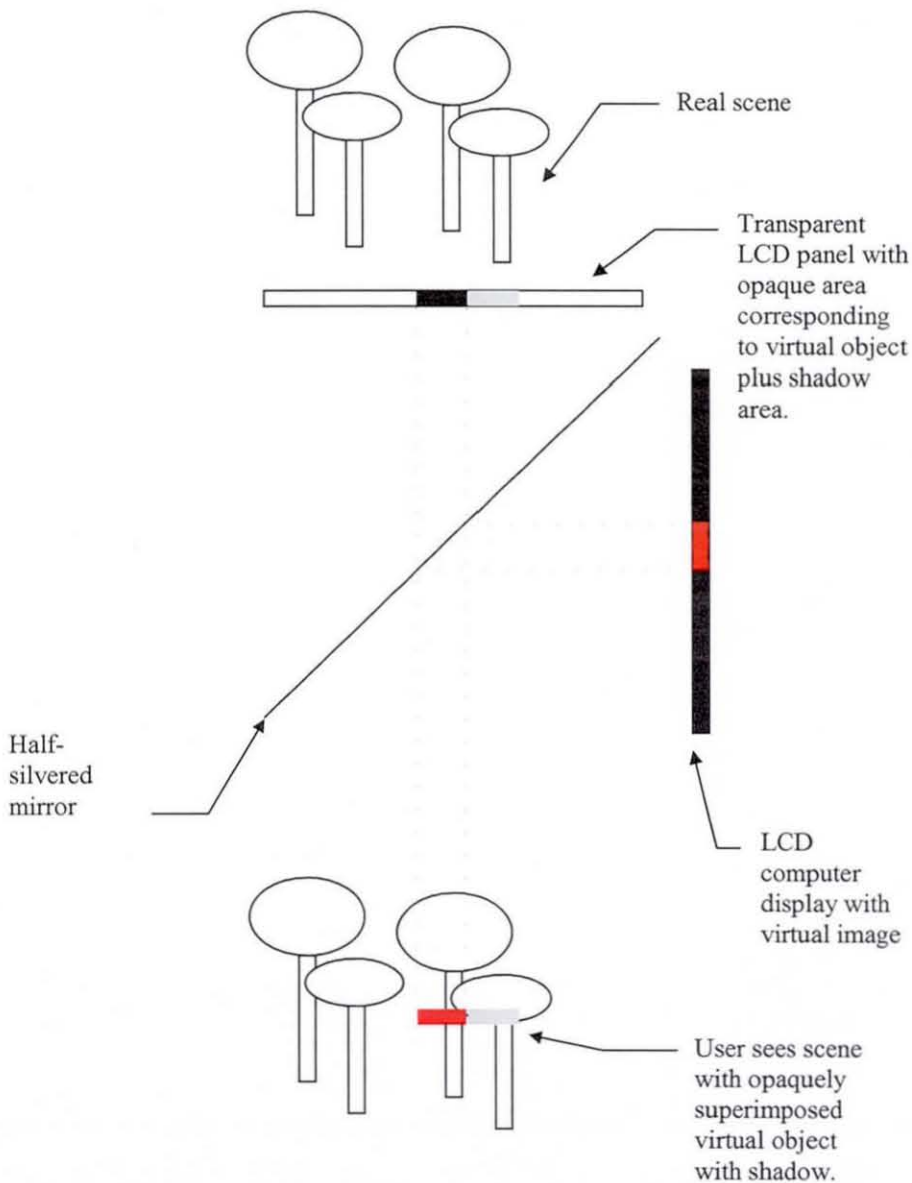
The transmission and projection approaches are explained and demonstrated in Sections 7.1 and 7.2 respectively. The reflection approach is discussed in Section 7.3.

### **7.1 Transmission approach**

Alpha-channel masking views of reality can be implemented by interposing an active transparent panel between the real scene and the viewer. Individual pixels of the transparent panel must be made more or less opaque, or coloured, as required in order to mask areas of the scene. The interposed panel could, for example, comprise a transparent LCD screen. Figure 7.1 illustrates this concept.

To explain the principle, it is helpful to make a number of simplifying assumptions. First, that the beam splitter has 50% reflectance and 50% transmittance with no absorption, and that there is also no inherent absorption by the active panel. Second, that all colours, including those in the real world, can be represented within a

common Red, Green and Blue gamut. Activating the panel allows subtractive control of intensity values seen from the real scene, reducing the intensity in selected areas as required. The reflected display panel effectively operates over the same intensity range but this time in an additive way so that selected areas of colour can be increased in intensity. The net result is that, at the cost of a 50% reduction in see-through light intensity, a significant degree of control over the image is obtained. (The arithmetic basis is explained fully in Chapter 6.) Figure 7.1 illustrates the basic display panel arrangement proposed.



(Figure 7.1 – Optical-see-through display employing transmissive alpha channel mask)

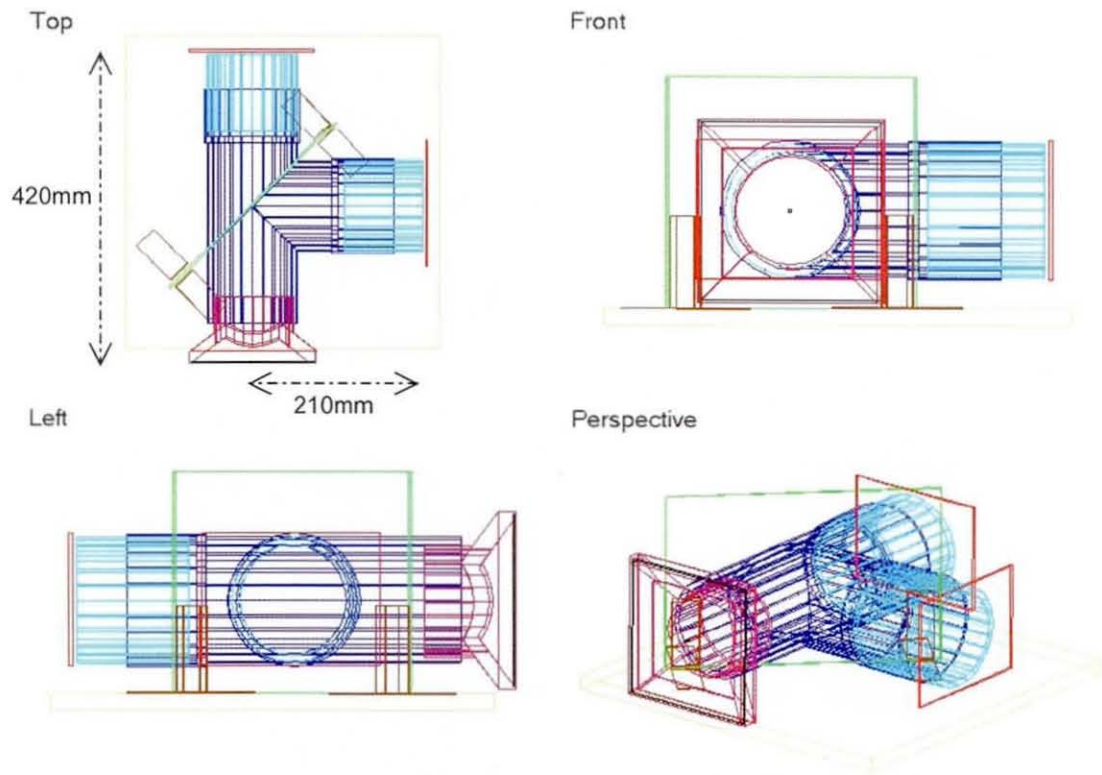


The following section describes the design and construction of a prototype monocular optical-see-through display arrangement capable of implementing a real-world transmissive alpha channel for displaying occlusion and shadow effects in augmented reality, based on the principle shown in figure 7.1.

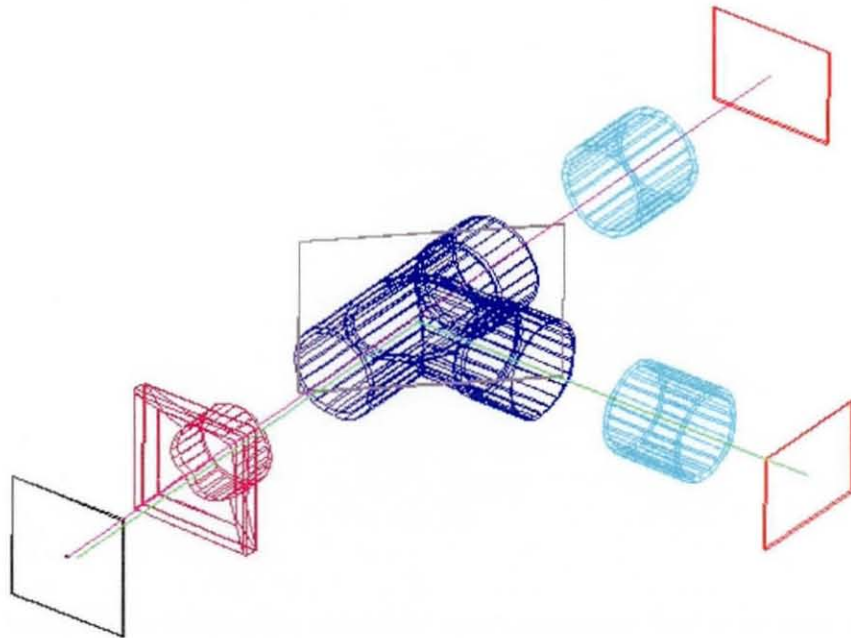
### **7.1.1 Building a prototype**

When considering the design of the prototype, and as wearability was not an issue at this stage, it was apparent the device built for the depth perception experiment in Chapter 5 would form a suitable basis.

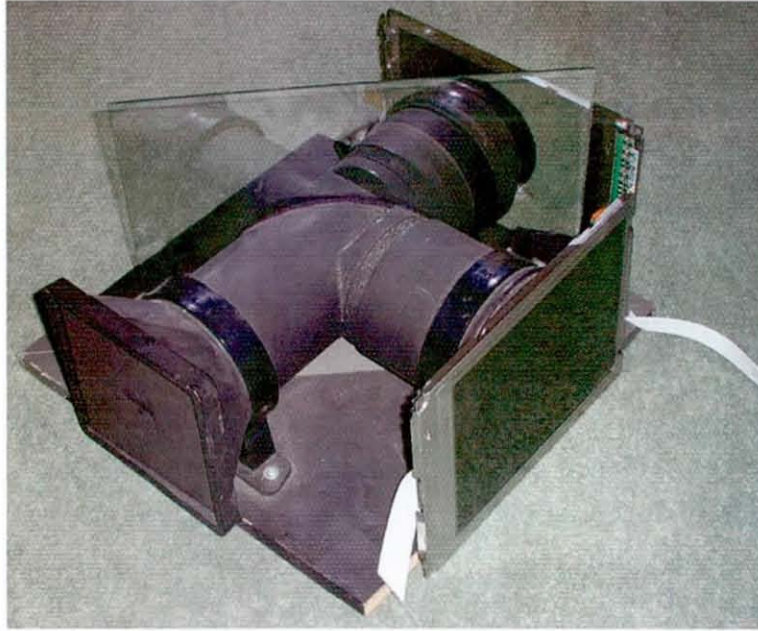
The display was constructed using drainpipe sections of 100mm internal diameter. The main body was formed from a T-junction that was cut across its centre section at 45 degrees to accommodate the beam-splitter. Separate pipe sections were fitted to the T-junction openings to provide for later pipe length adjustment. A rectangular drainage hopper was attached at the viewer end of the display with a UPVC panel cut to fit and drilled with a viewing hole. The complete assembly was attached using pipe brackets to a 400mm x 400mm baseboard. Adjustable angle brackets were also attached to the baseboard to hold the beam splitter and display panels. The whole assembly was painted matt black inside and out to reduce spurious reflections. All open pipe ends and brackets were lined with black felt to avoid damage to display surfaces and to provide a light-tight fit. The overlay display panel was an NEC NL8060AC26-11, 10.4 inch, 262144 colour TFT LCD panel, 800 x 600 pixels resolution with VGA driver electronics. The power supply was obtained by using a standard PC power supply unit. The masking panel was procured by cannibalising a nView Corporation overhead projection screen containing a transmissive TFT LCD panel of the same specification as the NEC panel. The driver circuitry and power supply from the projection screen were also utilised. Figure 7.2 shows plan, side, elevation and perspective views of the prototype. Figure 7.3 provides an exploded view to show the main component parts. Figure 7.4 is a photograph of the display with active LCD panels in place. Figure 7.5 shows a schematic diagram of the complete system.



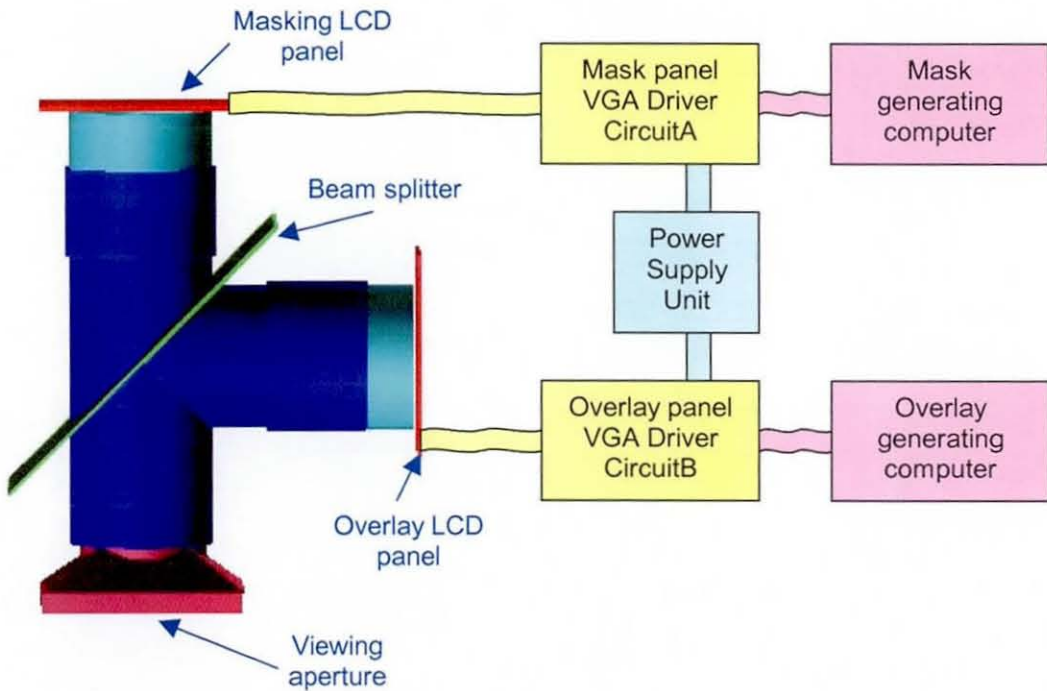
(Figure 7.2 – Display prototype; plan, side, elevation and perspective views)



(Figure 7.3 – Display prototype; exploded view)



(Figure 7.4 – Display prototype with display panels in place)



(Figure 7.5 – Schematic diagram of prototype system)



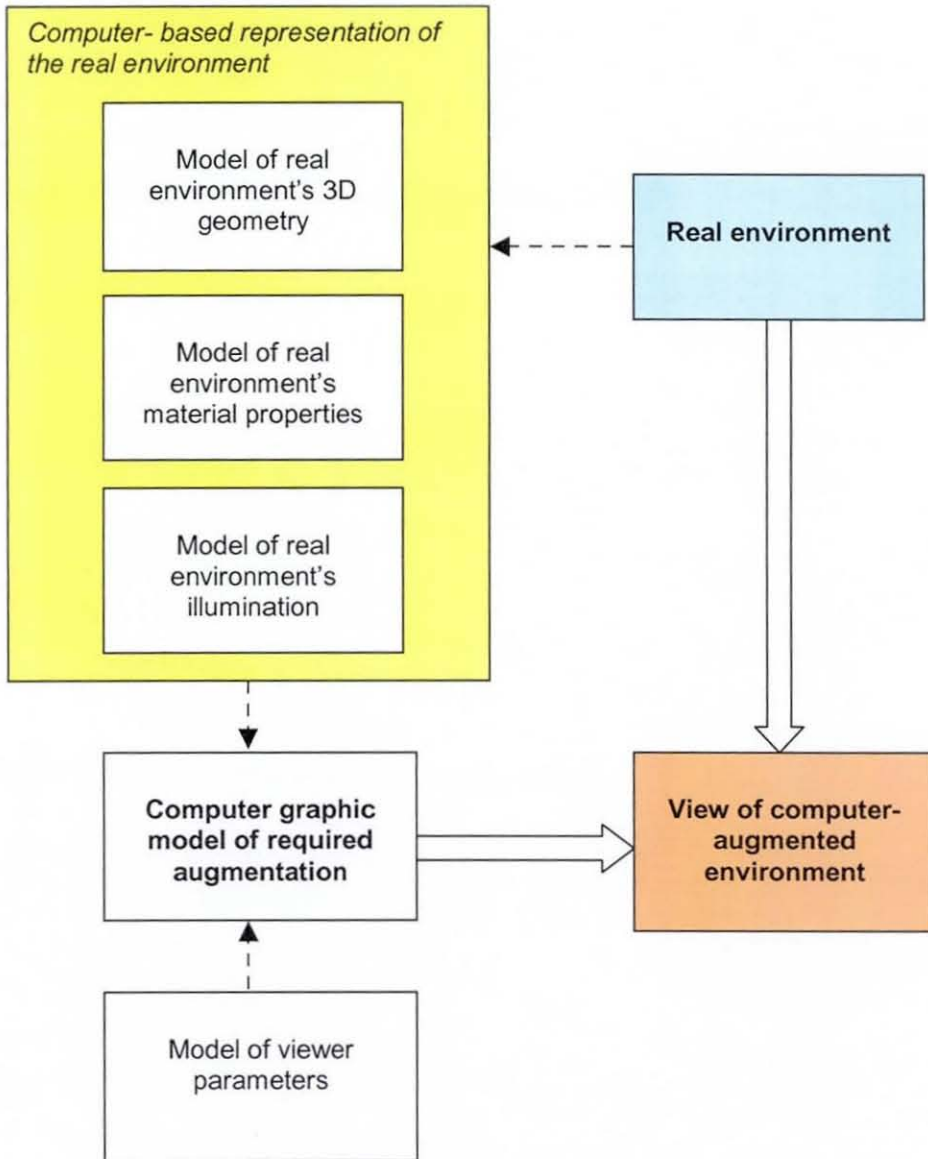
After construction of the prototype, tests were conducted to demonstrate the effectiveness of this approach to real-virtual compositing.

### **7.1.2 Compositing demonstration**

The purpose of these demonstrations was to verify the display's capacity to produce the following varieties of real-virtual augmentation:

- a virtual object occluded by a real,
- a real object occluded by a virtual,
- virtual shadows across a real surface,
- real shadows across a virtual surface,
- a virtual refracting object in front of an opaque real object,
- a real refracting object in front of an opaque virtual object,
- virtual illumination of a real scene.

For integration of real and virtual scenes in a way that can automatically take account of the above factors, it will ultimately be necessary to acquire sufficient information from the real environment with respect to geometry, material properties (eg surface reflectances) and its illumination. For accurate compositing, a computer-based representation of the real environment is needed. Ideally this would be comprised of the following environment models; its geometry, the material properties of all surfaces, and the illumination. This information would allow a photorealistic rendering of a real scene to be generated. Although this rendering is not required, coupled with the view parameters, these models provide the information needed to render a virtual object fully integrated within the real scene. Figure 7.6 shows the modelling components that would compose an ideal augmented reality system. A complete model of the real environment and the viewing model are used in the generation of the computer augmentation before virtual and the real worlds are composited.



(Figure 7.6 – Component models for an ideal augmented reality system)

Obtaining the necessary real world model information poses significant difficulties and these are explored in more detail in chapters 8 and 9.

In the meantime, if we assume that these problems will eventually be solved, we can envisage a perfect computer-augmented scene as appearing exactly as the real scene would if the augmented objects were themselves placed within it. Thus, convincing real-scene views before and after augmentation can be created photographically, the



augmented image providing an 'ideal' target for any computer-augmented reality system. This principle was used for initial testing of the prototype display.

### **Compositing test scenes**

For testing purposes, six scenes were created. To obtain accurate representations of augmented scenes without, at this stage, having to handle the complex issues of modelling reality and obtaining a common illumination model, each scene was photographed twice; once with and once without the augmentation. One photograph of each pair represented the original scene and the other, a photograph showing the ideal composite. For example, Test Scene 2 features a wooden toy lorry. This represents a real scene being viewed by an augmented reality system user. Its paired photograph is the original scene with the addition of a toy car, and this represents our target scene; that is, the representation of what we would expect in a perfect rendering of the original scene with augmented virtual car. Note that two small Xs have been added to each photograph for registration purposes.

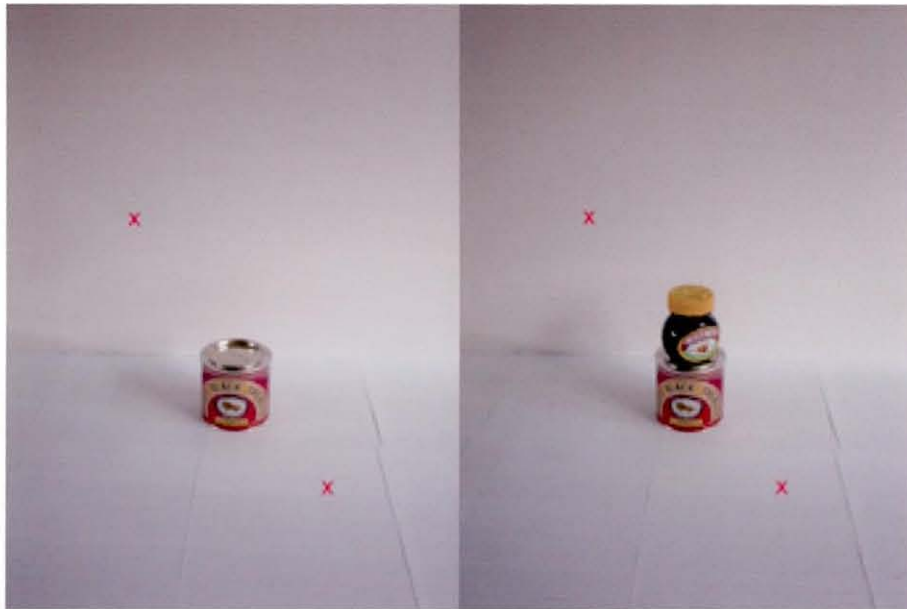
The actual test scenes are shown on the following pages as figures 7.7 to 7.12.

*Scene 1*

*Original:* A syrup tin against a white paper background

*Augmented:* A jar has been added such that it occludes part of the tin and background and casts a shadow on the background.

*Purpose:* This is primarily to demonstrate a virtual object occluding a real.



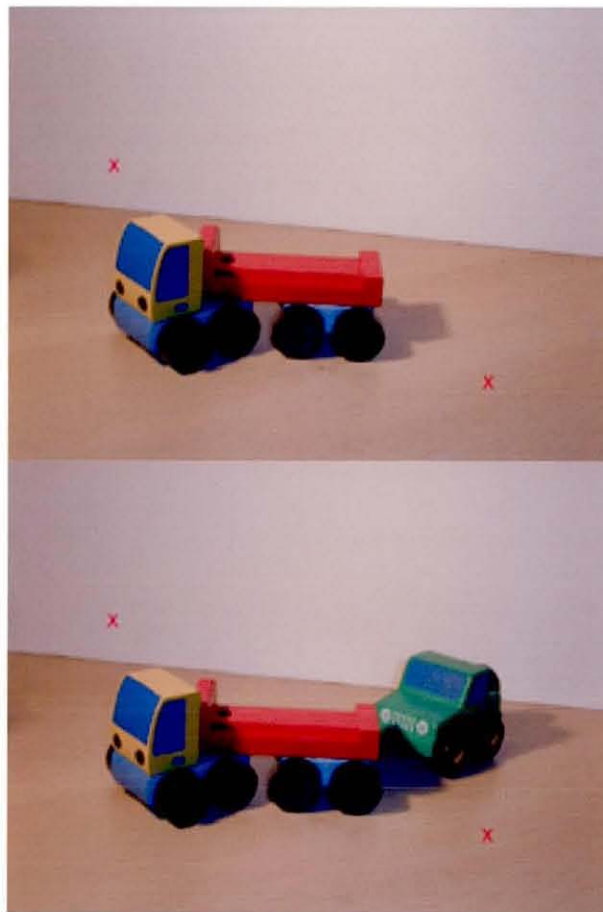
(Figure 7.7 – Test scene 1)

*Scene 2*

*Original:* A toy lorry on a wooden table

*Augmented:* A toy car has been added such that the lorry partially occludes the car and casts its shadow on it, while the car occludes and casts a shadow across part of the background.

*Purpose:* This was primarily to demonstrate a real shadow being cast across a virtual object, and a virtual object casting a shadow on a real surface.



(Figure 7.8 – Test scene 2)

*Scene 3*

*Original:* A bottle, a tin and a jar in front of a white paper background.

*Augmented:* A virtual box has been added such that is partially occluded by the existing objects.

*Purpose:* This was to demonstrate a virtual object being occluded by a real.



(Figure 7.9 – Test scene 3)

**Scene 4**

*Original:* A cylindrical container on a wooden table.

*Augmented:* A glass jug has been added such that it is partially in front of the container and it casts a shadow on the table.

*Purpose:* This was to demonstrate a refracting virtual object partially occluding a real object.



(Figure 7.10 – Test scene 4)



*Scene 5*

*Original:* A glass jug on a wooden table.

*Augmented:* A cylindrical container has been added such that it is partially behind the jug and it casts a shadow on the table.

*Purpose:* This was to demonstrate a virtual object being refracted by a real.



(Figure 7.11 – Test scene 5)

*Scene 6*

*Original:* A scene of assorted items with a lamp that is switched off lit by natural light.

*Augmented:* The scene illuminated using the lamp which has been switched on

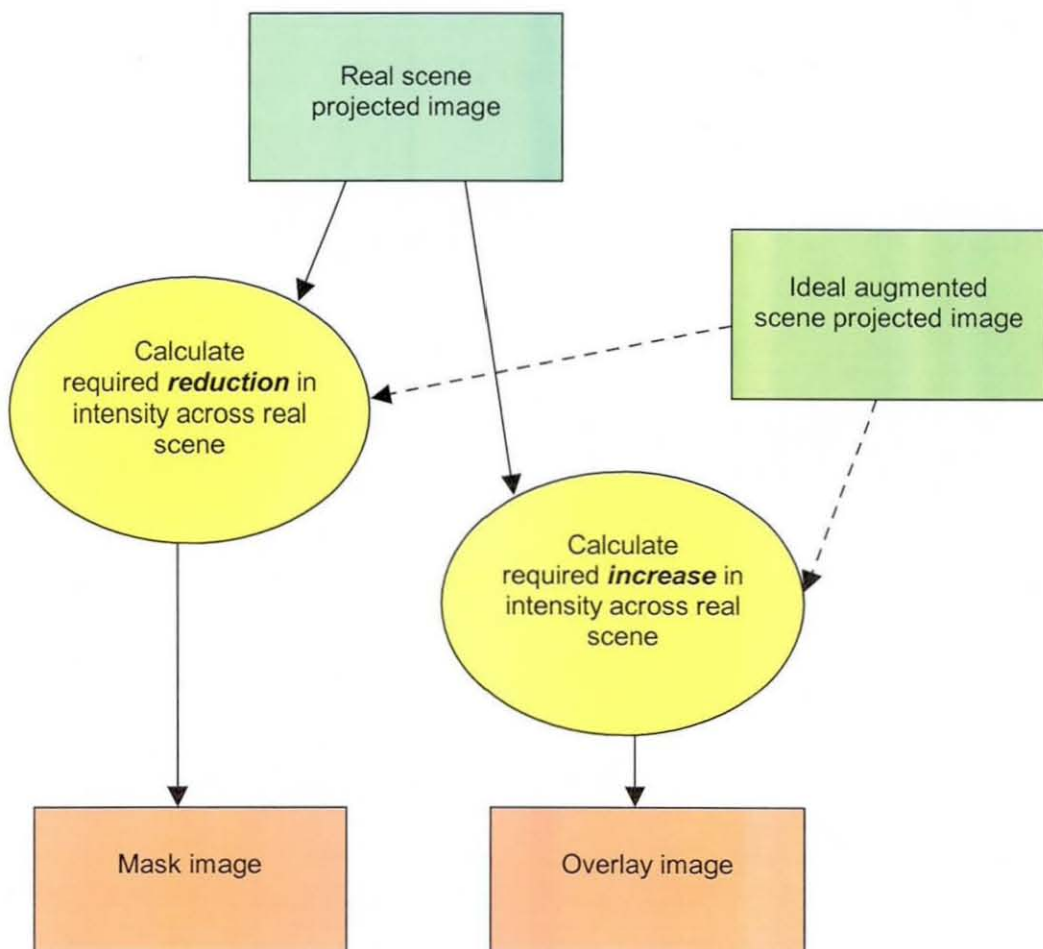
*Purpose:* This was to demonstrate virtual alteration to the lighting in a real scene.



(Figure 7.12 – Test scene 6)

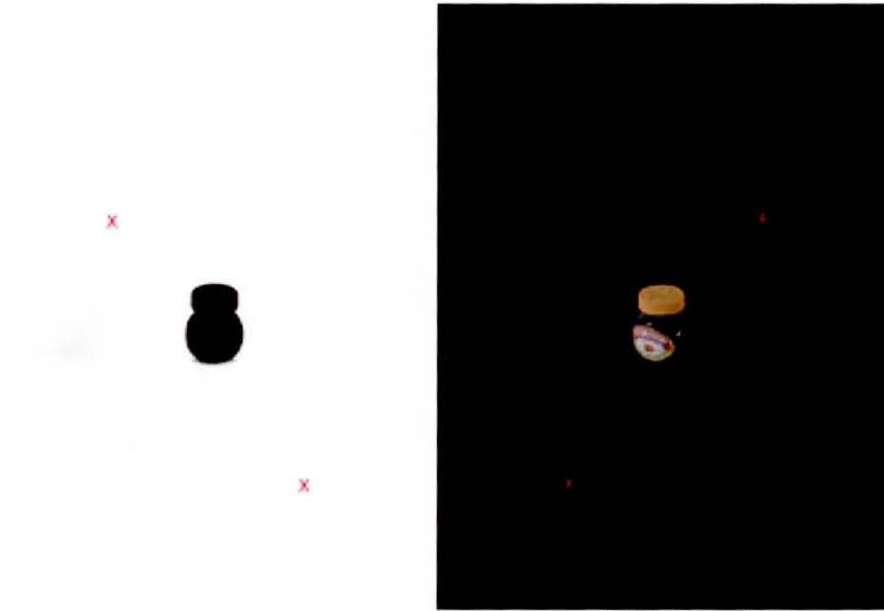
### Calculating the masks and overlays

For each original scene, the mask and overlay images required to produce the composite were calculated as detailed in Chapter 6. The differences between the real scene projection, which represents the user's view of the environment, and the required augmented view are used to calculate the necessary changes in intensity across the scene. As illustrated in figure 7.13, these changes determine the appropriate mask and overlay images.

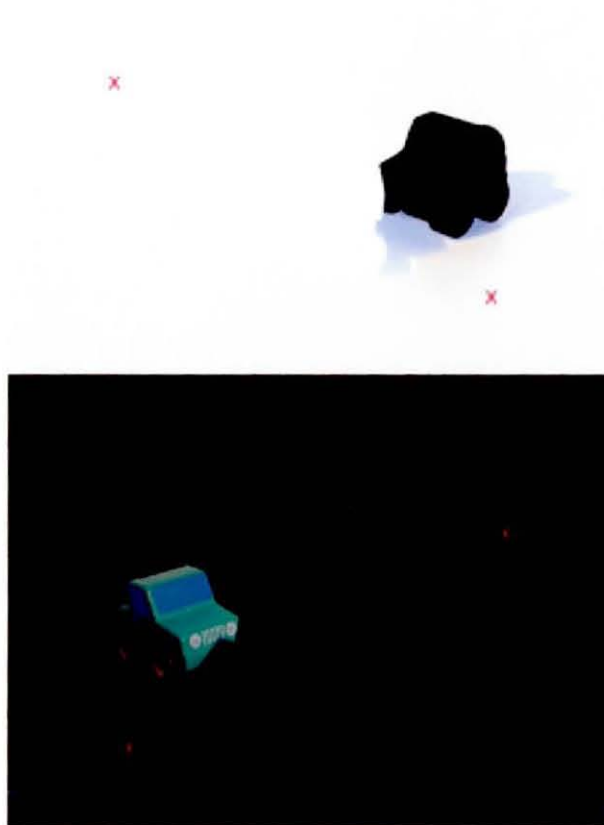


(Figure 7.13 – Schematic for production of mask and overlay)

For each scene, the masks and overlays actually determined are shown in figures 7.14 to 7.19.



(Figure 7.14 – Scene 1, mask and overlay)



(Figure 7.15 – Scene 2, mask and overlay)



(Figure 7.16 – Scene 3, mask and overlay)



(Figure 7.17 – Scene 4, mask and overlay)





(Figure 7.18 – Scene 5, mask and overlay)



(Figure 7.19 – Scene 6, mask and overlay)

### **Viewing**

Each original scene photograph was treated as a substitute for the original 3D scene and viewed through the display system with its calculated mask displayed on the transparent LCD panel. Its calculated overlay was displayed on the non-transmissive panel so as to be reflected in the half-silvered mirror. The positions of these display surfaces were carefully adjusted to register with the scene by aligning the registration marks.

### **Compositing results**

The resulting views through the display were photographed and videoed to record for each:

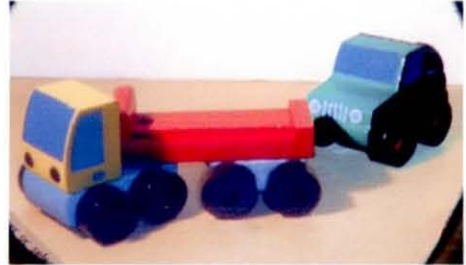
- the original scene viewed through the display without augmentation,
- the original scene with transmissive panel switched on to show alpha masking, and
- the original scene with both mask and overlay panel activated to reconstruct composite scene.

The resulting photographic images are shown in figure 7.20.

Scene 1



Scene 2



Scene 4



Scene 3



Scene 6



Scene 5



(Figure 7.20 – Augmented images)

### **7.1.3 Demonstration using a real scene**

This section describes demonstration of the display using a simple real 3D scene with animated augmentation. The scene used comprises a real pillar made from Lego™ bricks orbited by a virtual sphere.

#### **Scene preparation and modelling**

To minimise the issues relating to real-world geometry and lighting (for a fuller discussion see Chapters 8 and 9), a very simple scene was constructed so that its geometry and illumination was easy to model approximately. This scene consisted of a Lego™ pillar with 16 x 16mm cross-section and height of 100mm. It was lit by ambient lighting and also by an angle-poise lamp positioned above and to the left of the pillar.

Because it was necessary to calculate virtual shadows in relation to this real model, 3D viewing and rendering software was designed and written for the purpose. The coding was implemented using Borland Delphi Pascal and is described more fully in Chapter 9. The virtual object selected for the test was a sphere. This was chosen for its simplicity and ease of implementation as well as the fact that it could be used to suitably demonstrate occlusion and shadow effects. Ray tracing was used to render the sphere and to generate its shadow.

Whitted [WHIT80] proposed the first formulation of a recursive ray-tracing algorithm, while Glassner provides a detailed discussion of ray tracing [GLAS89]. The basic algorithm used in this demonstration is from Watt [WATT89].

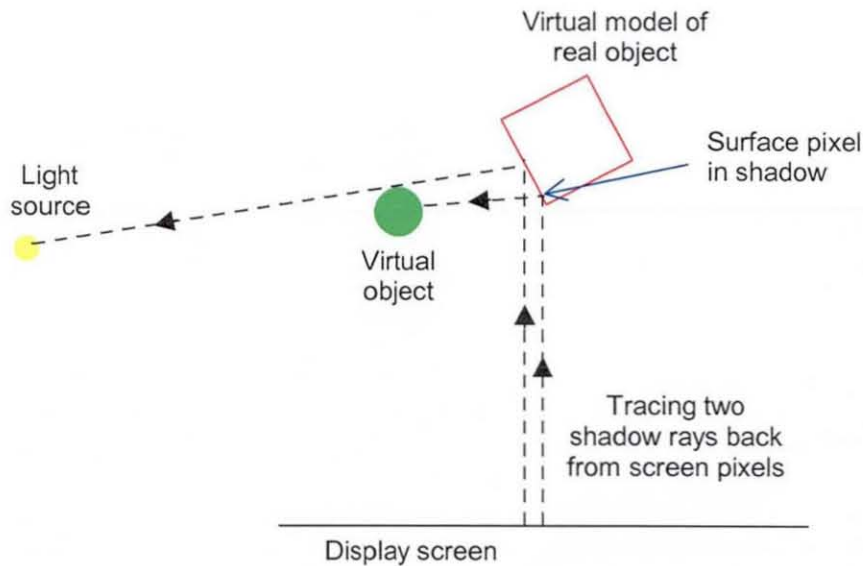
**Ray-trace algorithm:**

```

procedure RayTrace(start, direction: vectors; depth : integer;
                  var colour: colours);
var
  intersection_point, reflected_direction : vectors;
  local_colour, reflected_colour : colours;
begin
  {Intersect ray with all objects and find intersection point
  (if any) that is closest to the start of the ray}
  if {no intersection} then colour := {background colour}
  else begin
    local_colour := {contribution of local colour model at
                    intersection_point}
    if depth = maxDepth then reflected_colour := black
    else begin
      {Calculate direction of reflected ray}
      RayTrace(intersection_point, reflected_direction,
              depth + 1, reflected_colour);
    end;
    Combine (colour, local_colour, local_weight_for_surface,
            reflected_colour, reflected_weight_for_surface)
  end
end {RayTrace};

```

To implement shadows, feeler-rays track back to a point light source used to approximate the real world lighting as shown in figure 7.21, which is a plan view of the scene.

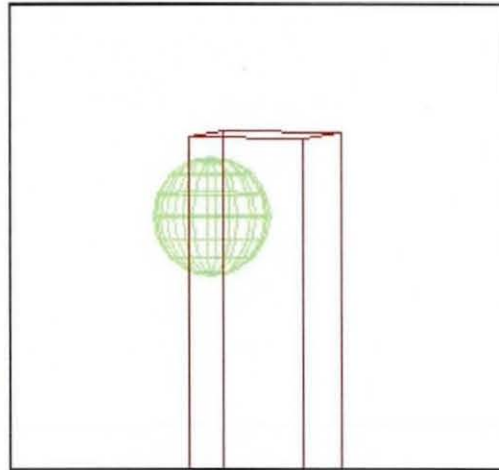


(Figure 7.21 – Shadow casting)



Unfortunately shadows produced by basic ray tracing do not exhibit natural looking soft edges or penumbrae. However, a modification devised by Cook, Porter and Carpenter [COOK84] produces more realistic blurring at shadow edges by perturbing some rays to follow paths other than that predicted by exact reflection angles.

An approximate virtual model of the pillar was produced by measuring the dimensions of the real pillar and replicating this with a simple graphic box. A virtual viewing position was selected such that the perspective of the virtual box closely matched that of the real pillar when viewed through the display. The computer-based graphic model included both this box and the sphere. (See figure 7.22.) The real pillar was 16 x 16mm in cross-section and 100mm high. Its nearest edge was 460mm from the viewer's eye. The virtual camera view angle was adjusted to provide a perspective view approximately matching the eye by viewing the virtual-pillar bounding box superimposed over the real pillar.



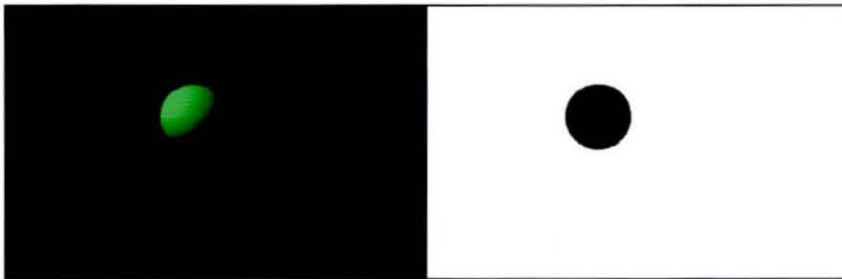
(Figure 7.22 – Geometric model of the scene)

To produce an animation of the virtual sphere rotating around the real pillar, the sphere was rendered in 24 positions along a circle of rotation to make it appear to orbit the pillar. For each position of the sphere, the ray tracer rendered only the sphere and the shadow as it would fall on the box; the box itself was not rendered. This produced a sequence of rendered images of the sphere and its shadow.

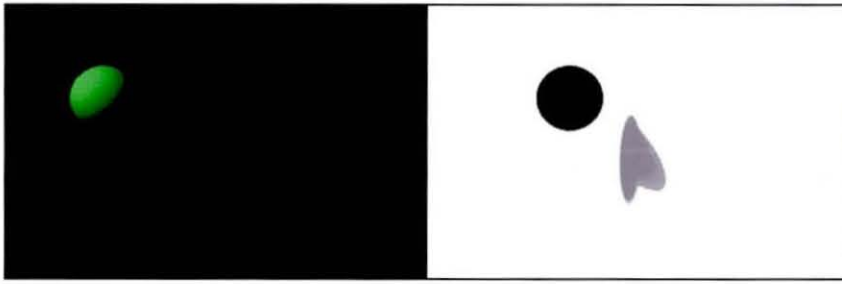
A commercial imaging-painting software package was used to create a holdout matte for each of the sphere images and to combine the holdout matte with the shadow to create the alpha-channel mask images. The sphere alone was used to create the overlay. The individual mask and overlay images were assembled as a short GIF animation.

For successful animation the mask and overlay images needed to be synchronised. Using two separate display panels made this problematic so a single computer was used to drive both panels with the screen buffer split into two halves; one for the mask and the other the overlay. Mask and overlay were registered with the real scene by eye and the animation run.

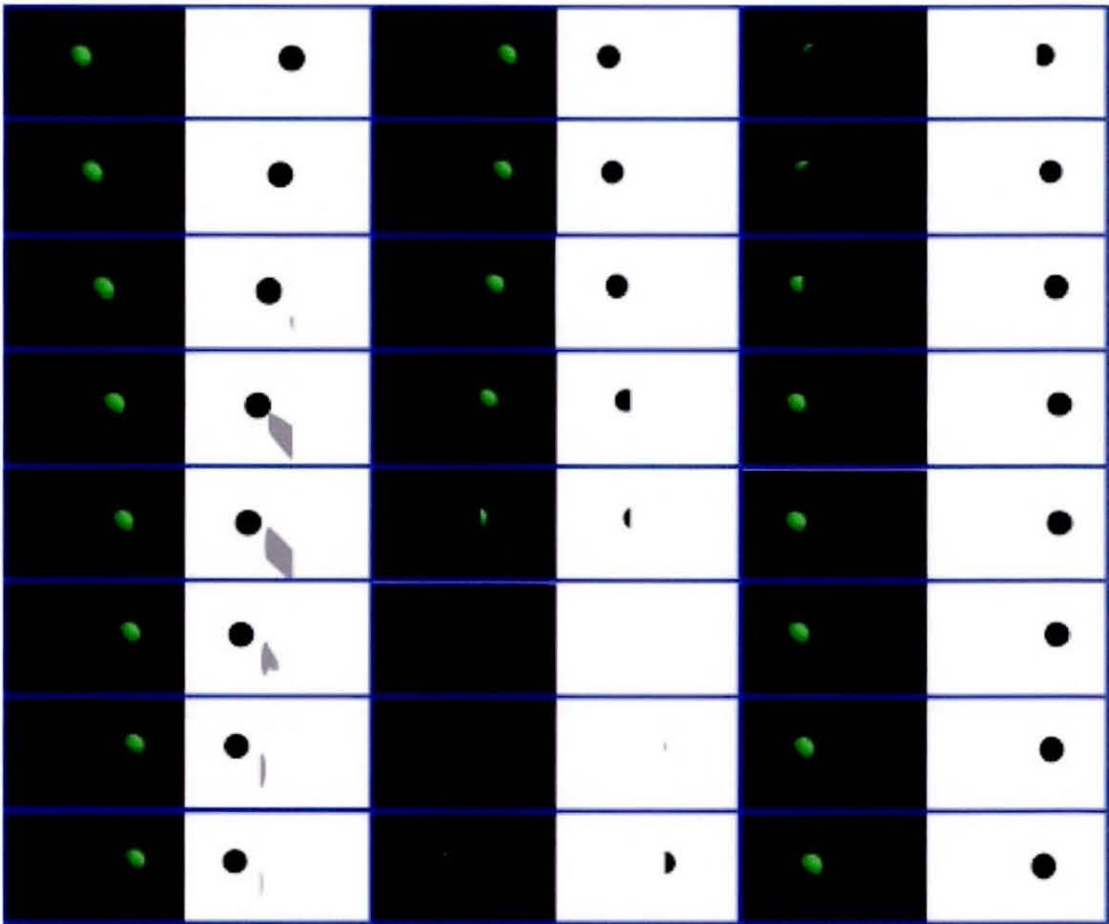
The illustrations in figures 7.23 and 7.24 show the mask and superimposition images used to create two of the frames from the animated sequence in which the virtual sphere orbits around the real Lego™ pillar. Figure 7.25 shows the complete animation sequence.



(Figure 7.23 - Superimposed reflection and transmission images)



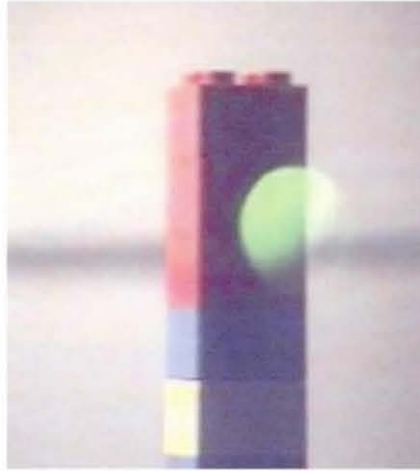
(Figure 7.24 - Superimposed reflection and transmission images)



(Figure 7.25 - Animation sequence; masks and overlays)

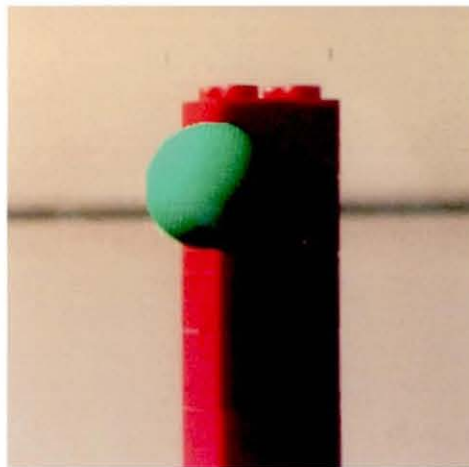
### Result

Figure 7.26 is a photograph of a Lego™ pillar taken through the prototype display without the alpha-channel mask in place. The ghost-like, transparent overlay characteristic of all current optical-see-through displays is evident.

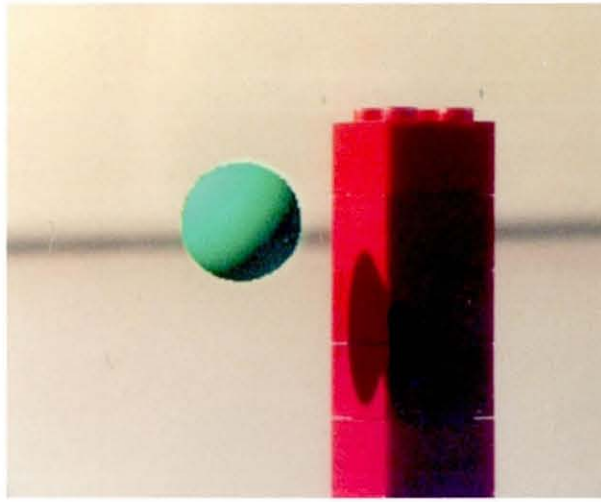


*(Figure 7.26 – Photograph through an optical-see-through display without masking)*

Figures 7.27 and 7.28 are photographs taken through the display with the alpha-channel masking active. Two positions in the orbit of the sphere are shown to demonstrate the occlusion and shadow effects achieved.



*(Figure 7.27 - Images composited with real-world view to show occlusion of a real background by a virtual object)*



(Figure 7.28 - Images composited with real-world view to show a virtual shadow on a real object)

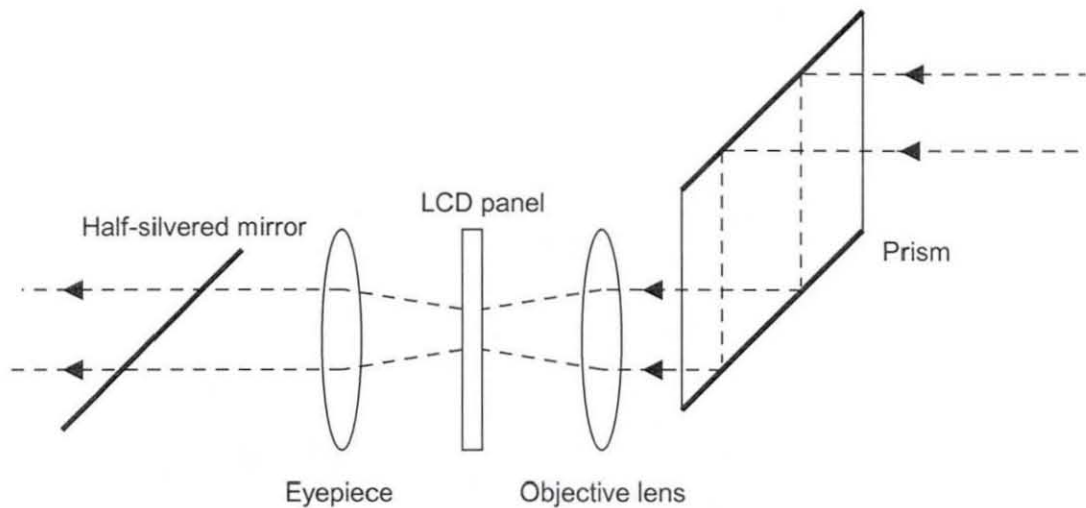
#### **7.1.4 Transmission masking discussion**

Using the transparent display element to create an opaque mask and the reflected element to display the superimposed graphic object, allows virtual entities to visually occlude a real background. The active transparent panel can also be used to generate areas of neutral density that reduce light received from selected areas of the real world enabling the simulation of virtual shadows within a real scene. Using a transparent active panel that is capable of displaying colour, it is possible to selectively filter areas of real world colour. Thus, the display arrangement is capable of visually simulating a range of virtual-real interactions. It is possible to produce colour-bleeding effects and even to display the effect of virtual light sources illuminating a real scene, as was demonstrated by Scene 6 in section 7.1.2.

In the demonstrations described above, the LCD panel was in very close proximity to the real scene objects. However, in most augmented reality applications the panel is likely to be further from the real scene and closer to the user. In this situation the mask and the scene cannot be simultaneously focused by the viewer. This accommodation problem needs to be corrected in a practical optical-see-through display.



Subsequent to the author's publication of some of the work described in this thesis [TATH99a] [TATH99b], Kiyokawa et al. [KIYO00] have addressed the focusing issue by placing two convex lenses, each with the same focal length, either side of the LCD panel and adding an erecting prism to correct inversion of the real scene. (See figure 7.29.)

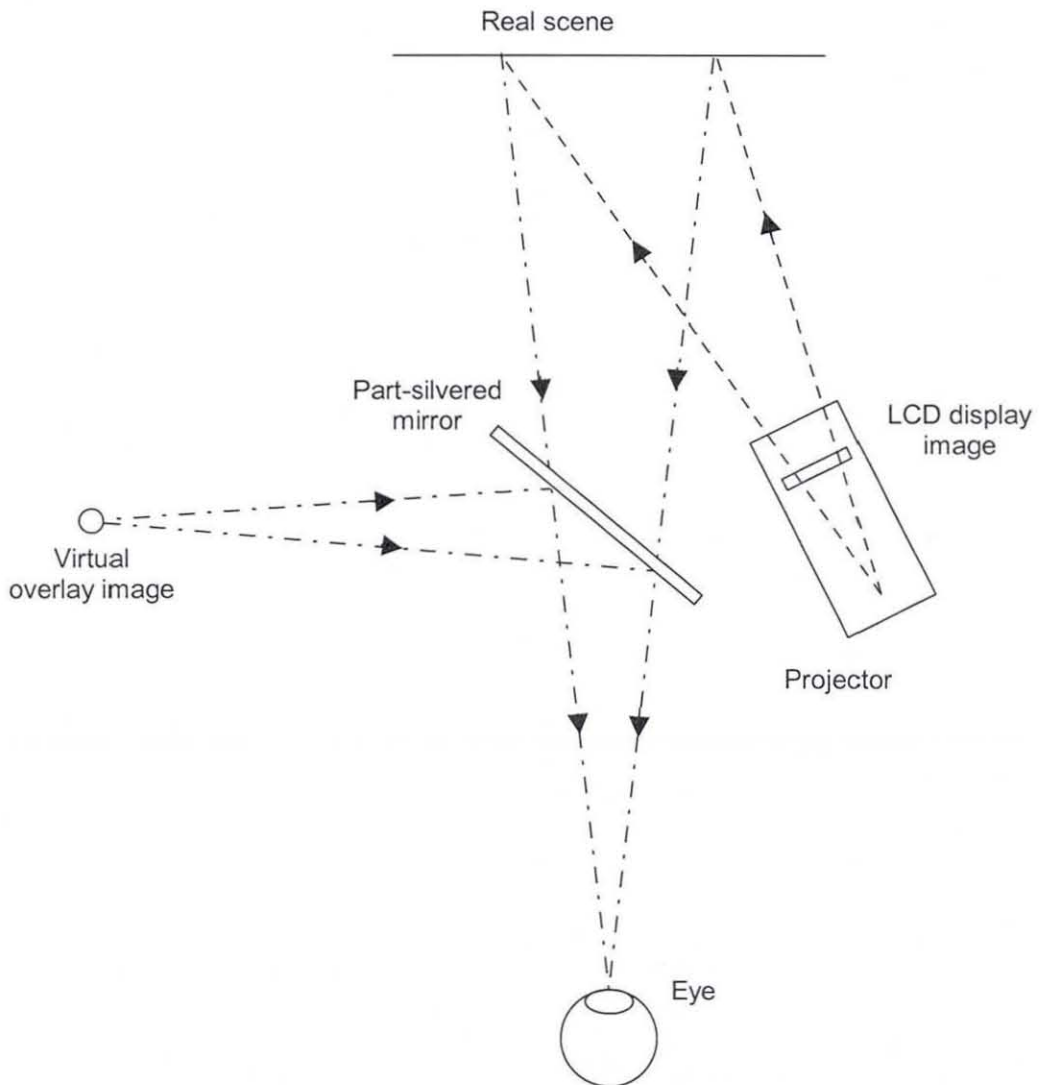


(Figure 7.29 – Optical arrangement to solve out-of-focus problem)

The transmissive filtering method based on the use of a LCD mask also suffers from the drawback that LCDs are polarisation-dependent. This means that 50% of the light intensity striking cannot be transmitted through the device, as the polariser filters it out. Light is also attenuated due to the transistors, gate and source lines in the LCD cells. Also, the liquid crystal material itself absorbs a portion of the light. Another disadvantage is the low fill factor, which is the percentage of the display surface actually used for display. This depends on the pixel size in relation to the inter-pixel spacing. Typically LCDs have up to 70% fill factor.

## 7.2 Projection approach

For some applications, where the scene is suitable and its illumination can be controlled, it is possible to achieve the real-world alpha channel effect by selectively illuminating the scene. For example, a digital projector could be used to light the scene so that areas that need to be masked are left unlit. This approach could be particularly apposite in medical applications where it is cumbersome for a surgeon to wear a head-mounted display or in situations where several viewers need to see an augmented scene.



(Figure 7.30 – Projection display arrangement)

For correct mask projection, the projector would need to be positioned directly along the user's line of sight but, as the projector would then obstruct the user's view or the user would obscure the projection, it is necessary to place the projector off centre so that it projects obliquely. See figure 7.30 for a possible display arrangement. The masked scene is viewed through a part-silvered mirror in which the graphic overlay is reflected. The oblique mask projections must be pre-warped to allow for oblique projection on a 3D scene. This process is described in Section 7.2.1.

### **7.2.1 *Projection masking demonstration***

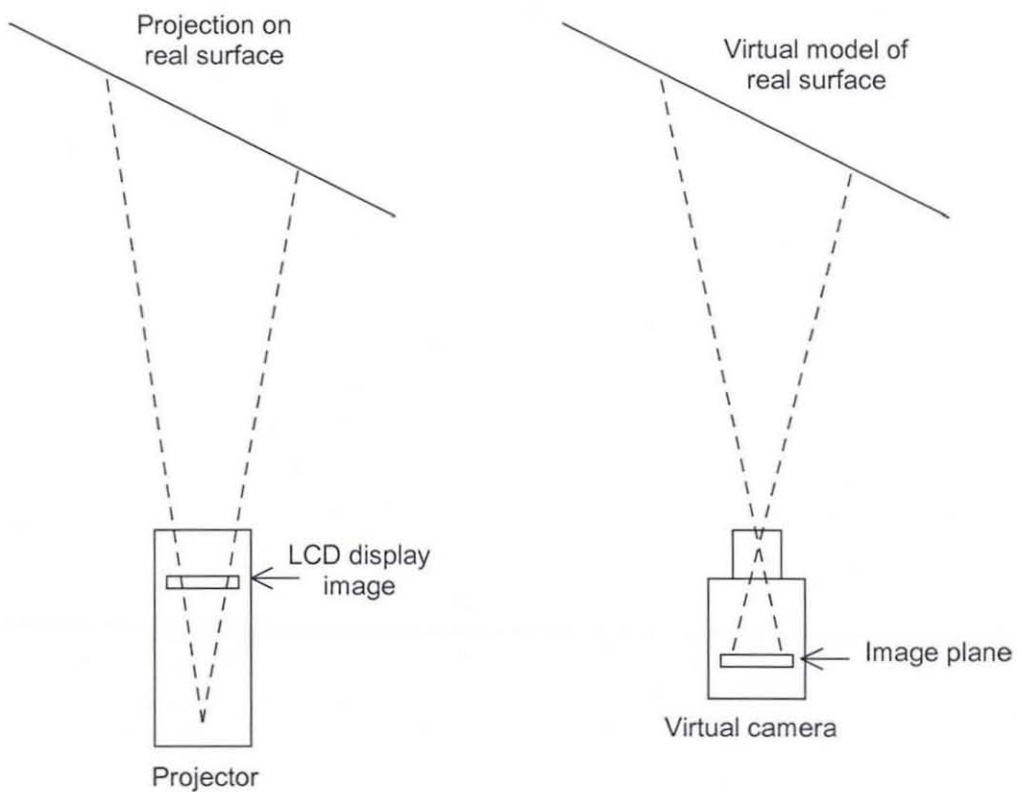
The author with the kind cooperation of the Royal College of Art and the Victoria & Albert Museum, South Kensington devised the demonstration described here.

The requirement was to display a 16<sup>th</sup> Century Polychrome Carving from the V&A collection. The artefact was approximately 1.5 x 1 metres, carved in wood, depicting St Christopher. It had once been elaborately decorated but, over the centuries, had lost much of its original finish. Chemical analysis had been carried out on the residual paint enabling determination of its original appearance. In some areas the decoration consisted of finely textured wax coating. The carving had already been 3D laser scanned so an accurate 3D computer model of the relief was available. This model had been coloured and textured appropriately so the polychrome could be viewed using a virtual reality head-mounted display. However, the museum was interested in being able to display the carving to the public in a way that would allow them to switch on or off different layers of virtual direction.

The monocular display arrangement illustrated in figure 7.30 was constructed with projection stands for the digital projector, and for a laptop computer, which provided the reflected overlay. A 50-50 beam splitter was held in position using a stand with adjustable clamps. Distances of image surfaces were adjusted manually until approximately correct registration was achieved. The overlay image was obtained from a projection of the 3D scanned model. The mask used was a simple holdout

matte produced from the overlay image. Because the mask was projected obliquely, it was pre-warped to counter the resultant distortion.

Correction for the oblique distortion is achieved for any projector position by capitalising on the fact that cameras and projectors with the same focal length are identical in geometric terms. A virtual camera in a virtual world containing the model of the St Christopher is placed in a position corresponding to the real projector. (See figure 7.31.) The image produced by the virtual camera then has the correct perspective view for the real projection. Thus, it is this image that is used to create the mask.



(Figure 7.31 – Correcting for oblique projection)

To ensure a fixed viewing position for the prototype display, the scene was viewed through a 35mm SLR camera viewfinder as illustrated in figure 7.32.

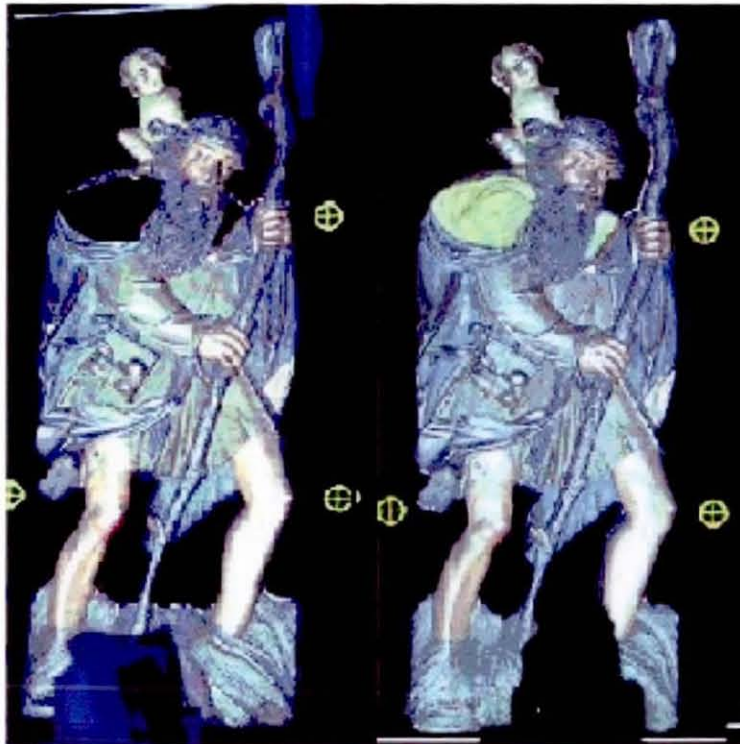




(Figure 7.32 – Viewing the scene through the projection prototype)

### Result

The images in figure 7.33 are photographs of the carving; first with the projected mask only and then with mask and overlay. In the first image, the alpha-channel projection acts as a holdout matte for St Christopher's cloak hood. The second image shows the virtual rendering overlaid in this area.

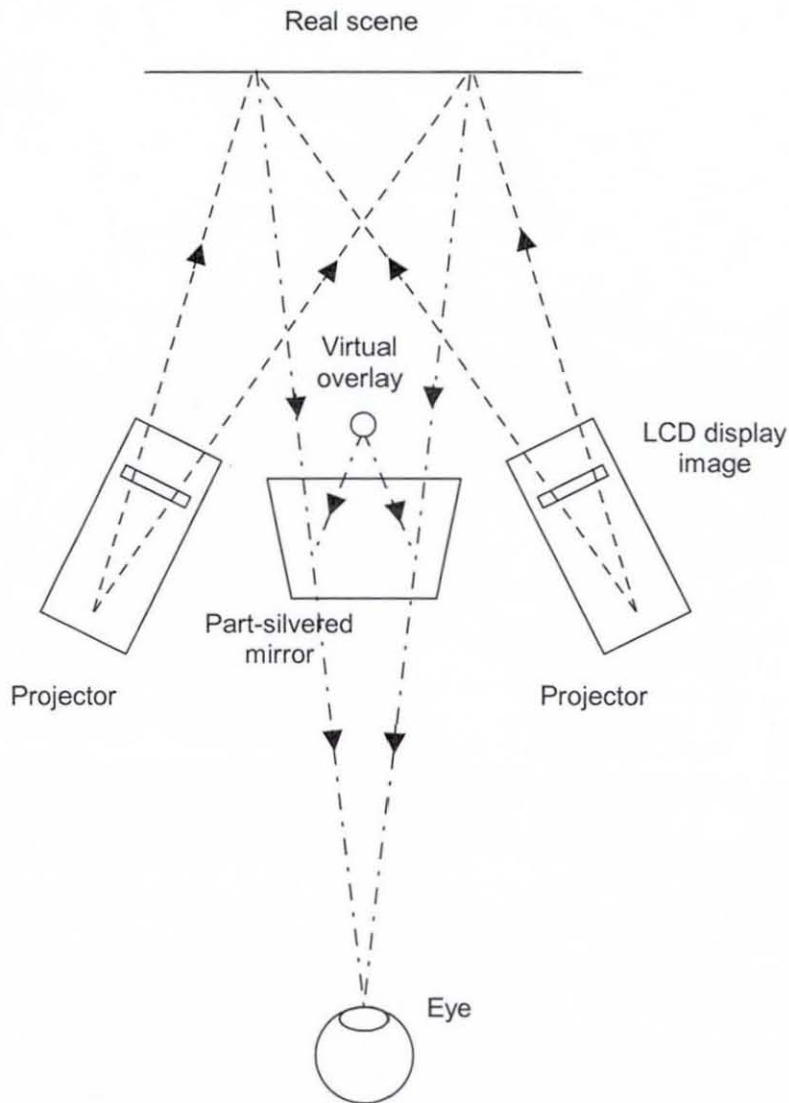


(Figure 7.33 – Scene with mask and with overlay)



### 7.2.2 Projection masking discussion

The polychrome carving had limited depth of relief. If this had not been the case, the oblique projection would have introduced undesirable shadows within the scene. To counter this problem, another oblique projection can be made from a second projector positioned to the other side of the user. Figure 7.34 shows this display set-up using two digital projectors, together providing the real-world masking. In this arrangement, the overlay display surface can be placed above or below the plane of the user's eyes.



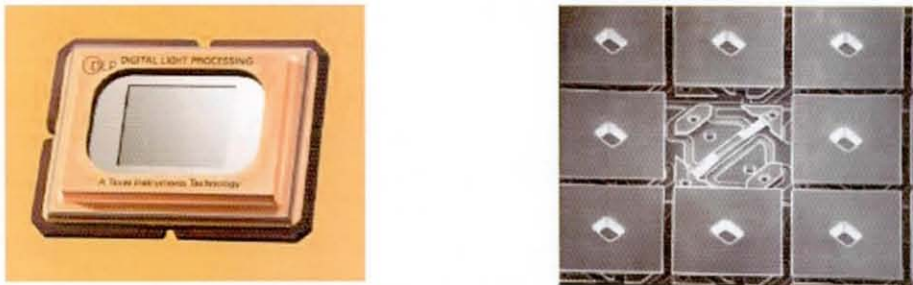
(Figure 7.34 – Display arrangement using two masking projectors)

The projective approach has the advantage that users need not be encumbered by a head-mounted display. However, it does require an environment in which the lighting can be completely controlled by the projector(s) and the viewer's position must be constrained to ensure the overlay is in register.

### 7.3 Reflection approach

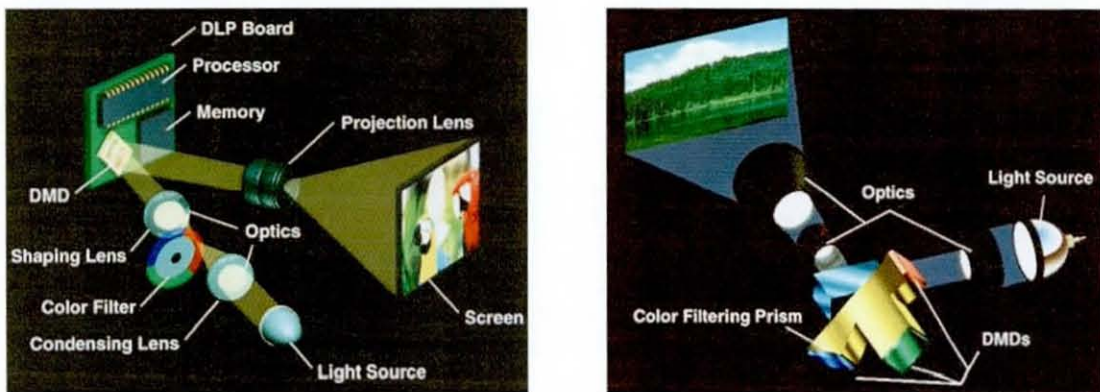
A possible alternative that overcomes the attenuation disadvantage of using a transmissive panel and the controlled environment requirement of projection is to employ a reflection approach based on the use of Microelectromechanical System (MEMS) technologies. One such suitable MEMS device is the Digital Micromirror Device (DMD™) developed by Texas Instruments [HORN89][SAMP93].

A DMD is, in effect, a semiconductor light switch comprising thousands of tiny, square,  $16 \times 16 \mu\text{m}$  mirrors, fabricated on hinges as a substructure over a static random access memory. (See figure 7.35.) The hinges allow the mirrors to tilt between two states,  $+10$  degrees for "on" or  $-10$  degrees for "off" thus allowing each mirror to switch a pixel of light. When the mirrors are not operating they rest at  $0$  degrees. A digital video or graphic signal sent to the DMD causes each pixel of information to be mapped directly to its own mirror. By electrically addressing the memory cell below the mirror it can be electrostatically tilted to the on or off position. The mirrors are capable of switching on and off at a frequency greater than  $1000\text{Hz}$ . This rapid modulation allows digital gray scale and color reproduction.



(Figure 7.35 – Digital Micromirror Device  
and magnified view with one mirror removed. Texas Instruments )

The usual current use of these devices is in DLP (Digital Light Processing) projectors. In these systems, after passing through condensing optics and a color filter system, the light from the projection lamp is directed at the DMD. When the mirrors are in the on position, they reflect light through the projection lens and onto the screen to form a digital, square-pixel projected image. Projection configurations can comprise a single DMD, or utilise a separate DMD for each RGB colour, as shown in figure 7.36.



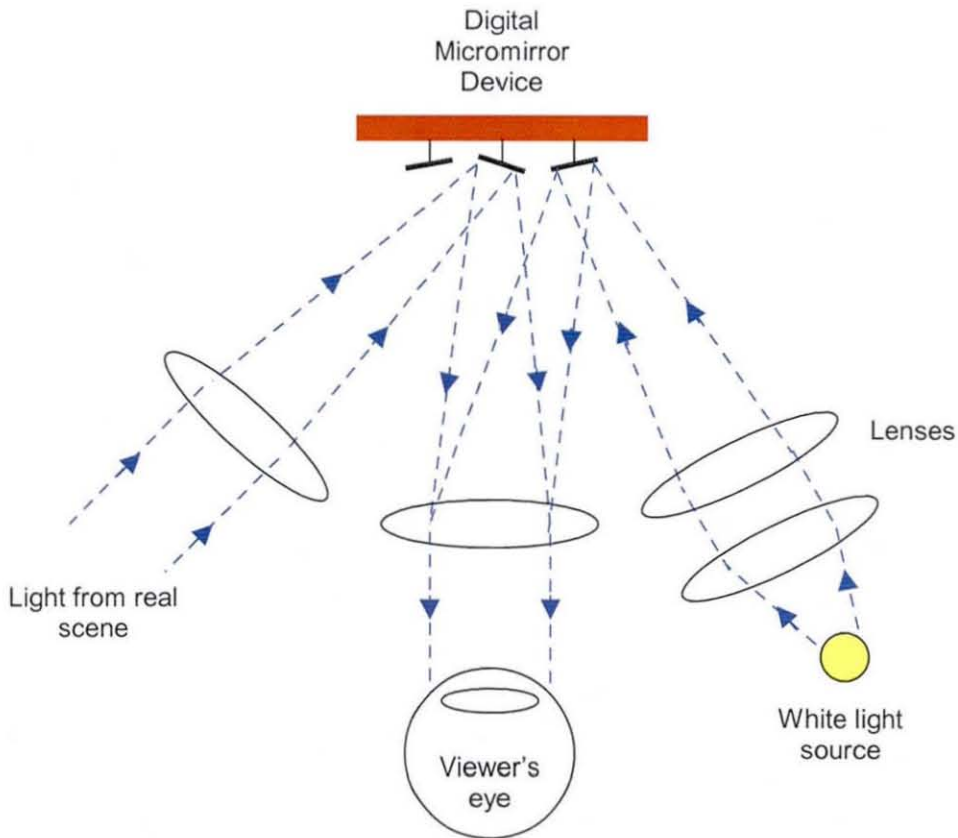
(Figure 7.36 – Left: Projection using a single DMD

Right: Projection using a separate DMD for each of RGB, Texas Instruments)

DMDs are reflective devices so attenuate light intensity less than a transmissive LCD. The square mirrors on DMDs are  $16 \mu\text{m}^2$ , separated by  $1 \mu\text{m}$  gaps, giving a fill factor of up to 90%. This high fill factor gives a higher perceived resolution than a LCD with the same pixel density.

Figure 7.37 shows a possible display arrangement for augmented reality. With the mirrors tilted in one direction (say  $-10^\circ$ ), they reflect the real scene to the user's eye. At  $+10^\circ$  light from a white light source is directed towards the viewer and at  $0^\circ$ , no light is reflected. Modulating mirror angles between  $0^\circ$  and  $+10^\circ$  provides the virtual image. Inclusion of a filter wheel or use of three DMDs would allow a coloured virtual image.





(Figure 7.37 – Augmented reality display arrangement using a DMD)

Alpha-channel adjustment of the real-world view becomes possible by modulating between  $0^\circ$  and  $-10^\circ$ . The longer the relative period spent at  $0^\circ$ , the less the contribution of light from the real scene that reaches the user.

The author first published the concept of using DMDs in this way in 1999 [TATH99a][TATH99b]. In 2002, a working prototype constructed at Osaka University [UCHI02] demonstrated the efficacy of this method, although this work is unpublished and reflects the computer graphic image from a conventional display rather than using the modulated approach described above. The consequence of this is that individual DMD mirrors either reflect the real scene or the virtual without the

possibility of blending between them. Figure 7.38 shows two original scenes used; one representing reality and the other a virtual object. The lower images show the composites produced.



(Figure 7.38 – Image splitting using DMD. Osaka University)

A wearable, stereoscopic augmented reality display using DMDs remains a considerable challenge.



#### **7.4 Discussion**

In this chapter three novel approaches to implementing a real-world alpha channel for optical-see-through displays have been described and their efficacy is demonstrated. These include introducing an active transmission filter capable of masking light intensity from the real scene, selectively lighting the real scene using projection, and merging real and virtual scenes using Digital Micromirror Devices.

The transmissive approach using LCD suffers from the drawbacks that light is attenuated significantly through LCD panels and the low fill factor further attenuates and pixelates the real-world view. These problems can be overcome by projecting the mask image onto the real scene. However, this can only be done in situations where there is no environmental lighting other than the projected illumination of the scene. An approach using MEMs such as DMDs overcomes these disadvantages although practical, wearable stereoscopic displays of this type will be challenging to fabricate.

Optical-see-through displays capable of displaying appropriate occlusions and shadows are an essential ingredient in the convincing visual integration of real and virtual scenes. However, their use to full advantage is predicated on being able to acquire a sufficiently accurate model of the world that is to be augmented. Chapters 8 and 9 consider this issue and propose a new strategy for acquiring a common illumination model from a real scene.

## **8.0 Scene reconstruction**

The ability to convincingly integrate real and virtual scenes depends largely on the accuracy of the real-world model that can be obtained. This chapter reviews the extent to which a fully accurate model is achievable and the strategies that can be employed to this end.

A real environment may already have a corresponding computer model to represent it; for example, an environment that has been constructed from a virtual reality representation. However, in such cases, there is certain to be differences between the model and the 'as-built' reality, which make it unlikely that design models could be used for augmented reality without correction.

In some circumstances it may be possible to measure and 'hand-construct' a model to represent a real scene although this is a painstaking process, depending on the accuracy of the representation required. Consequently, a number of strategies have and are being developed aimed at automating the reconstruction of 3D scenes. Broadly, techniques for recovering scene geometry can be classified as passive or active. The former attempt to draw inferences based on the visible radiation that already exists in the scene; typically using photographs. Active techniques involve some sort of interference with the scene itself, either by contact using a 3D position-sensing probe or by radiating energy into the scene.

### **8.1 Passive sensing**

Images from passive sensors can be used to reconstruct elements of scene geometry. Passive sensing is usually via camera-produced photographs or using video sequences. Some approaches for model construction from photographs require a high level of manual intervention; for example, edge extraction to facilitate semi-automatic reconstruction with the user fitting a wire frame model to corresponding image projections. Typical of more automated approaches are attempts to acquire 3D shape information from shading, stereo and/or motion.

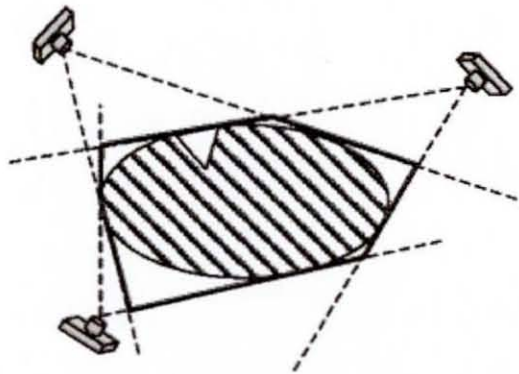
### 8.1.1 *Shape from shading*

In theory, given a 2D projection and assuming a light reflection model, it is possible to deduce something of the 3D shape of an object by variation in shade across its surface. Horn pioneered work in this area in the early 70s [HORN70]. Some assumptions are made to make this approach tractable, such as; the assumption that the surface reflectance map is known, and that the surface is illuminated by an infinitely distant point light source. Zhang et al. [ZHAN94] provide an overview of techniques along with a comparison and performance analysis.

Unfortunately, in practice, the simplified light models used in computer graphics are only approximate simulations of the behaviour of real light, and shape from shading is not yet a viable method for complete scene reconstruction.

### 8.1.2 *Shape from silhouettes*

Early attempts at reconstructing 3D models from photographs were based on the estimation of shape from silhouettes. A 3D object projects a 2D silhouette of its outline in a photograph and photographs from different viewpoints will represent different contours of the object. Taken together these can be used to define an intersection volume called a 'visual hull' [LAUR99], which can be assumed to fully enclose the object.



(Figure 8.1 – 3D object shown hatched with its inferred visual hull in bold)



The accuracy of the visual hull improves as the number of images is increased. However, silhouettes hide object concavities, which cannot be resolved from any viewpoint. This shape-from-silhouette approach has the advantage of being fairly straightforward if only an approximate model is required. However, reconstruction can only be carried out for small solid objects that can be viewed from multiple directions. This cannot be successfully applied to complete scenes, although it may be a suitable approach if only a rough model of real-world objects is required.

### ***8.1.3 Shape from stereo or motion***

If two slightly displaced images of a scene are available, it may be feasible to find correspondences between the images and hence compute depth using a process of triangulation. The underlying principle is similar to that of stereo vision, which relies on the slight disparity between two views of a scene to estimate depth. Although two views are sufficient to compute some depth information, more camera views may be employed. Establishing correspondence between views is a fundamental difficulty with this approach. One strategy is to correlate pixels of similar intensity. The actual position of a point in 3D space can be assumed to be somewhere along the ray connecting the optical centre of a camera and the projected position of the corresponding point in the camera image. Tracking projections across multiple images allows fairly accurate location of the point in 3D using triangulation [HART97].

To automatically extract 3D scene geometry from multiple 2D images, it is necessary to extract and match corresponding image features. If this can be done for a sufficient number of points and lines then estimates of 3D locations of the features and can be deduced. These techniques can be applied to multiple static views of a scene or from motion sequences in which a camera is moved around the scene [MAYB93].

#### *8.1.4 Shape from photo-consistency*

If we assume that under conditions of constant illumination and Lambertian reflectance, scene surface points appear to be the same colour from all different views, photo-consistency can be used as a basis for determining 3D geometry [SEIT97]. This is a volumetric approach in which the principle is to encompass the whole scene in a volume subdivided into voxels. Voxels are assumed to be opaque until they fail the photo consistency test and are labelled transparent, or vice-versa. Testing for photo-consistency entails projecting every voxel centroid onto each of the images from which it is visible and thresholding the colour variance of the corresponding image pixels. The difficulty that arises is that several model solutions could be consistent with the projected images used. [KUTU99] have addressed this issue and devised a 'space carving' algorithm that yields a unique reconstruction solution called the 'photo hull'. It is the union of all photo-consistent subsets of the scene.

To improve efficiency, various further improvements to the original 'space carving' algorithm have been proposed, including generalised voxel colouring (GVC) [CULB00] in which a layered depth image data structure is utilised so that a linked list of voxels, sorted in depth order, is associated with each image pixel over which they project.

Slabaugh et al. [SLAB00] have used GVC for large-scale scene reconstruction by dividing the voxel space into an interior space where reconstruction of foreground surfaces takes place, and outer space which is used to model the background scene. This approach appears to give good results for outdoor scenes but is not applicable to indoor environments where wall structures restrict visibility.

Volumetric approaches to rendering large scenes become difficult as the number of voxels required becomes excessive. Future multi-resolution approaches may help mitigate this problem.



## 8.2 Active sensing

Besides using a physical probe to plot the contours of an object in 3D space, which is a slow painstaking process, other techniques involve emitting some form of radiation into the scene. 3D data can be obtained using X-rays (as in Computer Tomography), microwave radar, or sonar range finding. However, suitable active approaches for augmented reality are more likely to be optical, of which there are two main categories; laser range scanning and triangulation of structured light. In the former case, laser light is transmitted into the scene and the reflections detected. Measuring the difference in phase or time of flight can be used to determine the distance to the scene's object surfaces. This is a fairly established technique and Nitzan et al. [NITZ77] provide an early description of distance measurement based on the phase difference between an emitted and received beam of laser light. Typically, in the case of structured-light triangulation, a line of light is projected across the scene and viewed by camera. Because the separation of the light source and camera is known, depth can be computed by measuring displacement along the projected image of the light line.

While laser scanners do produce an accurate depth map within their field of view, a single scan will usually contain holes due to occlusion. Also near objects will be sampled at higher resolution than far. Therefore, to obtain a complete model, the scene must be scanned from multiple locations then the results 'stitched' together. Determining the best set of scanning locations to minimise number of scans required is a difficult problem and itself the subject of some research effort [eg FLE100]. The need to stitch separate scans together also leads to the problem of accurately aligning the range images with one another. One approach is to control the motion between views so that it can be suitably calibrated. This can be done by attaching the sensor to a robot arm or by placing the object to be scanned on a turntable. These strategies are clearly not viable for complete scenes. Alternatively, correspondences between points in overlapping images can be used to compute rotation and translation vectors to determine the relationships between different views. This requires fairly robust

methods of feature detection, and is far from trivial to automate [GREE99][ROTH99][ZHAN99].

Registration algorithms are generally based on the Iterative Closest Point (ICP) method [BESL92], which consists of two main steps. Given two sets of points, it is first necessary to identify pairs of candidate points for likely correspondence and, secondly, to compute a transformation that minimises, in a least-squares sense, the distance between two sets. This process is repeated until some convergence criterion is met.

Once registered, the range images must be integrated to produce a single 3D model. This can be accomplished by combining the 3D mesh representations of overlapping regions [eg RUT194] or by using volumetric-based methods [eg CURL96].

Although acquisition, registration and integration of multiple-scan images is successful for modelling small objects, reconstruction of large 3D scenes leads to increasing computational complexity. An attempt to reconstruct a complete scene has been presented by El-Hakim et al. [HAK198] who describe a system consisting of eight CCD cameras and a laser range scanner mounted on a mobile platform. The scanner is used to produce a range map for each of the eight camera images. Features corresponding to discontinuities in the image and range maps are extracted automatically. However, the matching does require manual intervention. Correspondences between 2D and 3D features are used to register the range data, which are then integrated using a volumetric method.

Approaches to reducing the large volume of data yielded by scanning real scenes include plane fitting algorithms which try to fit planes to 3D coordinate measurements [SEQU99].

### 8.3 Texture acquisition

The simplest approach to texture acquisition is to map photographs onto the geometry of the scene. To map a photograph we need to know the intrinsic camera parameters and its position and orientation in the world. Unfortunately, real cameras introduce distortions and do not behave the same as the perfect projection from a pinhole camera generally assumed by computer models. However, it is possible to model the distortions created by real lenses [TSA187].

The next task is to register the 2D images to the 3D scene geometry. Specifying point correspondences between images and geometry enables camera pose to be determined. For example, Corrêa et al [CORR02] describe the interactive selection of correspondences to obtain good camera calibration allowing pictures to be taken from any position.

Once all parameters have been determined, pixel colours from the camera images can be mapped to the 3D model.

### 8.4 Discussion

Unfortunately, none of the currently available techniques provides a way of acquiring fully automatic and accurate reconstructions of real-world scenes. The most promising method at present is to use laser range scanning, which is more accurate and less sensitive to scene illumination and environmental conditions than reconstruction approaches using photographs alone. However, scanning equipment is expensive compared to low-cost, high-resolution digital cameras and, as not all augmented reality applications require a complete model of the real environment, passive techniques may sometimes suffice. The usefulness of being able to acquire 3D models from real scenes extends beyond the requirement for augmented reality, making this a very active focus of ongoing research and an area in which future significant advances are likely.

## 9.0 Common illumination

As discussed in section 3.2.2, one of the most significant hurdles to convincing integration of virtual entities within a real scene is achieving an appropriate unification of illumination such that real light sources illuminate virtual objects and vice-versa. The situation is particularly complex as light radiates from surface to surface within an environment until an equilibrium is reached, effectively making every point on every surface a potential source of reflected, if not emitted, energy that can potentially influence every other point in the environment. Introducing new objects upsets this balance necessitating a global reconfiguration of light interactions.

If the original scene has been fully modelled *a priori* then the position and quality of all surfaces and light sources are known and conventional rendering techniques may be used. However, whereas standard renderers seek to colour surfaces according to the position of light sources and the juxtaposition with other surfaces, the requirements for obtaining a common illumination model are basically the reverse; that is, to determine the position and intensity of light sources and unseen reflective surfaces that will account for the illumination and consequent shading of known surfaces.

This chapter presents a new approach for inferring a common illumination model from a restricted view of a real scene. It requires that scene geometry and surface reflectance properties are known in advance but offers the advantage over existing methods that the positions of the real light sources do not need to be known or mapped beforehand using a physical probe. Sections 9.1 and 9.2 provide an overview of illumination modelling and radiosity rendering. In section 9.3, previous methods for solving the common illumination problem are reviewed. The theory and practice of the new approach are explained in sections 9.4 and 9.6. Section 9.5 describes a 3D viewing and rendering implementation developed for this work. A concluding discussion is provided in section 9.7.

## 9.1 Illumination modelling

This section provides an overview of illumination models typically used in computer graphics.

A 'local illumination model' considers only the contribution to the shading of a surface coming from direct illumination by a light source. A 'global illumination model' also takes into account the effect of light reflected from one surface to another. A 'global model' is therefore more physically correct and produces more realistic rendering. However, it is more computationally expensive.

The Phong model [PHON75] is a commonly used local illumination model for computer graphics. This model can only deal with point light sources and has three components:

$$I = \text{ambient} + \text{diffuse} + \text{specular}$$

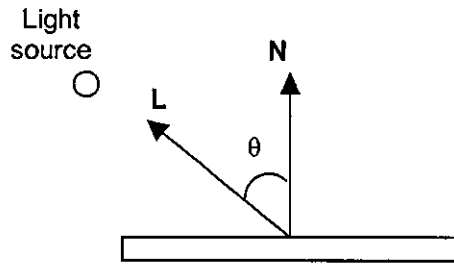
The ambient term is included to allow for some global control of brightness in a scene. This is assumed to have a constant intensity throughout the scene. Each surface, depending on its physical properties, has a coefficient of ambient reflectance, which measures what fraction of this light is reflected from the surface. Hence for an individual surface the intensity of ambient light reflected is:

$$I = k_a I_a$$

where  $I_a$  is the constant intensity of the ambient light, and  $k_a$  is the coefficient of reflection of the surface.

Diffuse reflection is usually assumed to be Lambertian [LAMB60]; that is, a perfectly diffuse reflecting surface that scatters light equally in all directions. Thus the intensity at a point on a surface as perceived by the viewer does not depend on the position of the viewer. When only diffuse light is considered, surfaces will appear dull or matt.





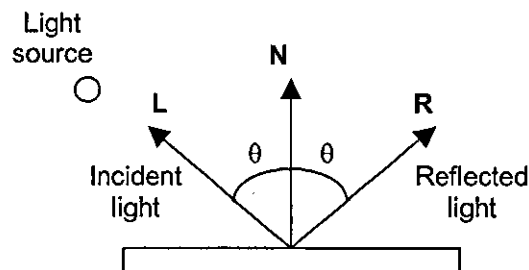
(Figure 9.1 – Lambertian reflection)

In figure 9.1,  $\mathbf{N}$  is the unit surface normal and  $\mathbf{L}$  is a unit vector in the direction from the point to the light source. Lambert's cosine law gives the intensity due to diffuse reflection:

$$\begin{aligned} I_d &= I_i k_d \cos\theta \\ &= I_i k_d (\mathbf{L} \cdot \mathbf{N}) \end{aligned}$$

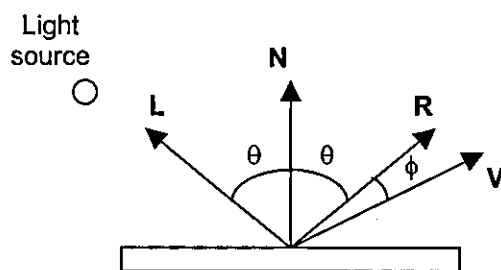
where  $I_i$  is the intensity of the incident light,  $\theta$  is the angle between the surface normal at the point and the line from the point to the light source, and  $k_d$  is the coefficient of diffuse reflection or the material and is dependent on the wavelength of the incident light.

Specular reflection (figure 9.2) is caused by the mirror-like properties of a surface. A perfect mirror will reflect light arriving at the surface at an angle of incidence,  $\theta$ , to the normal at a reflected angle of  $\theta$  to the normal in the same plane as the normal and the incident light. This means that only a viewer looking along the reflected ray  $\mathbf{R}$  will see the reflected light.



(Figure 9.2 – Specular reflection)

In practice no surface is a perfect mirror and there will be a certain amount of light scattered around the reflected direction. The reflected light is therefore seen over an area of the surface as a highlight. (See figure 9.3.)



(Figure 9.3 – Specular highlight)

In practice the distribution function for specularly reflected light is a complex function of  $\phi$ , the angle between the reflected ray  $\mathbf{R}$  and the viewing direction  $\mathbf{V}$ . Phong modelled this distribution empirically as:

$$\cos^n \phi$$

For a glossy surface,  $n$  is set to a high value and for a matt surface to a low value. Hence we obtain the complete basic reflection model:

$$I_s = I_i k_s (\mathbf{R} \cdot \mathbf{V})^n$$

where  $k_s$  is the coefficient of Specular Reflection.

Combining the various models and assuming the Phong illumination model gives:

$$I = I_a k_a + I_i k_d (\mathbf{N} \cdot \mathbf{L}) + I_i k_s (\mathbf{R} \cdot \mathbf{V})^n \quad \dots \text{eqn 9.1}$$

where each of  $k_a$ ,  $k_d$ , and  $k_s$  are parameters which are associated with specific surfaces and take values between 0 and 1.

More recently, physics-based illumination models have been applied to produce more realistic results. These use a more accurate model of reflectance in the form of a “Bidirectional Reflectance Distribution Function”. The BRDF describes how much light is reflected when light makes contact with a particular material. It depends on the viewer and light source position relative to the surface normal and tangent. As different wavelengths of light may be absorbed, reflected, and transmitted to varying degrees, BRDF is also a function of wavelength. For most real materials, that are not homogeneous, BRDF will also vary across a surface.

A general BRDF can be written using functional notation as;

$$\text{BRDF}_\lambda(\theta_i, \phi_i, \theta_o, \phi_o, u, v)$$

where  $\lambda$  is used to indicate that the BRDF depends on the wavelength;  $\theta_i, \phi_i$  represent the incoming light direction in spherical coordinates;  $\theta_o, \phi_o$  represent the outgoing reflected direction in spherical coordinates, and  $u$  and  $v$  represent the surface position parameterised in texture space.

## 9.2 Radiosity rendering

The radiosity method provides a solution to the global illumination problem. It stems from a formulation first developed by Siegel and Howell [SIEG84] to account for heat transfer between elements in furnaces and on spacecraft. The approach is based on using the principle of conservation of energy or energy equilibrium to find a solution accounting for the radiosity of all surfaces within an enclosure. The original ideas were applied to computer graphic rendering by Goral, Torrance, Greenberg and Battaile [GORA84] who used the method to model diffuse light phenomena such as colour bleeding, shading within shadow envelopes, and penumbræ along shadow boundaries, all of which effects are beyond the capacity of former conventional rendering techniques such as ray tracing.

The method is based upon a simple model of energy transfer. At each surface in a scene, the amount of energy that is given off is comprised of the energy that the surface emits itself plus the amount of energy that is reflected off the surface. The amount of reflected energy can be further characterised as the product of the amount of energy incident on the surface and the reflectivity of the surface.

The radiosity method is based on the following two principles.

1. Energy is conserved at each point on a surface; that is, light energy incident upon a surface and generated in the surface must balance the light energy emitted from the surface or absorbed by it in the form of heat.
2. Each point on a surface serves as a source of light for illuminating every other surface element in the scene within its line of sight or indirectly through reflection from other surfaces.

To simplify the computation involved, basic radiosity algorithms ignore spectral reflections and they also assume that only perfect Lambertian reflection takes place.

The two principles of energy balance can be expressed by the radiosity relationship:

$$B_i = E_i + \rho_i \int_{\text{environment}} B_j F_{ij} \quad \dots \text{eqn 9.2}$$

where

$B_i$  is the total rate of energy leaving surface  $i$  (ie its radiosity). Units are those of power density (energy per unit time per unit area,  $\text{Wm}^{-2}$ ).

$E_i$  is the rate of light emission from surface  $i$ .

$\rho_i$  is the reflectivity of surface  $i$ .

$B_j$  is the total rate of energy leaving surface  $j$ .

$F_{ij}$  is called the form factor. It is the fraction of light energy leaving surface  $j$  which strikes surface  $i$ .

Written in terms of a finite number of elements,  $n$ , equation 9.2 becomes:

$$B_i = E_i + \rho_i \sum_{j=1}^n B_j F_{ij} \quad \dots \text{eqn 9.3}$$

There is such an equation for each surface patch in an enclosure and the complete environment produces a set of  $n$  simultaneous equations which can be written conveniently in matrix form  $\mathbf{Ax} = \mathbf{b}$  as:

$$\begin{bmatrix} 1 - \rho_1 F_{11} & -\rho_1 F_{12} & \dots & -\rho_1 F_{1n} \\ -\rho_2 F_{21} & 1 - \rho_2 F_{22} & \dots & -\rho_2 F_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -\rho_n F_{n1} & -\rho_n F_{n2} & \dots & 1 - \rho_n F_{nn} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ B_n \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \cdot \\ \cdot \\ E_n \end{bmatrix} \quad \dots \text{eqn 9.4}$$

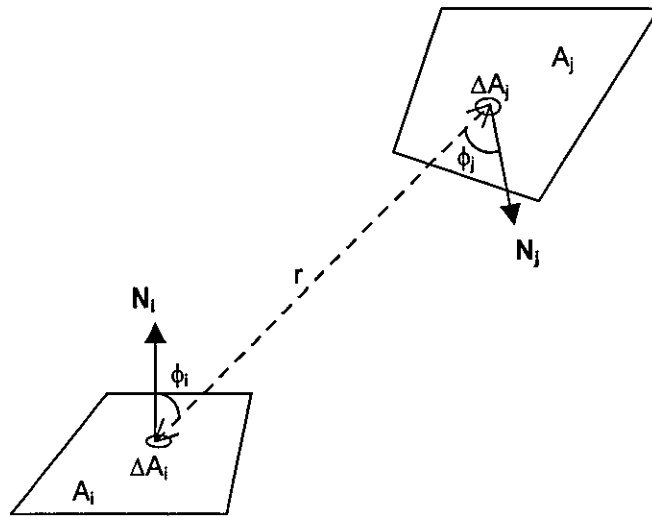
The  $E$  values are all zero except for surface patches that emit light by conversion from some other form of energy and therefore represent the input illumination for the system. The reflectivities,  $\rho_i$ , are known or can be calculated, and form factors  $F_{ij}$  are a function of the geometry of the system. In normal circumstances, for synthetic graphics rendering, matrix  $\mathbf{A}$  is known, as is the vector  $\mathbf{b}$ . The  $\mathbf{x}$  vector represents the unknown radiosities and can be determined easily using a straightforward iterative scheme such as the Gauss-Seidel method. (See Strang [STRA88] for description.)

In practice, each source and reflective patch is assigned an RGB value based on its material properties. Radiosity calculations are carried out to determine patch intensity in terms of three radiosities,  $B_R$ ,  $B_G$ ,  $B_B$ , each representing one of the RGB colour components.



The form factor in the above equations is defined as the fraction of energy that leaves surface  $i$  and reaches surface  $j$ . It is therefore a proportion in the range 0 to 1. The form factor values depend on the size, distance and orientation of surface patches relative to one another and can be calculated from scene geometry. Where one surface patch is hidden from another due to some occluding object, the corresponding form factor is set to zero, thus providing a mechanism for hidden surface removal.

Fig 9.4 shows the parameters involved in computing form factor  $F_{ij}$ .



(Figure 9.4 – Radiosity form-factor parameters)

Assuming planar patches  $A_i$  and  $A_j$  with dimensions within an order of magnitude of the distance between them, it is necessary to integrate the effects of differential areas  $\Delta A_i$  and  $\Delta A_j$  on each other.

Cohen and Greenberg [COHE85] describe the normalised differential form factor:

$$F_{\Delta A_i \Delta A_j} = (\cos \phi_i \cos \phi_j) / \pi r^2$$

Integrating over  $A_j$  determines the influence of the whole of patch  $A_j$  on the differential element  $\Delta A_i$ :

$$F_{\Delta A_i A_j} = \int_{A_j} [(\cos \phi_i \cos \phi_j) / \pi r^2] dA_j$$

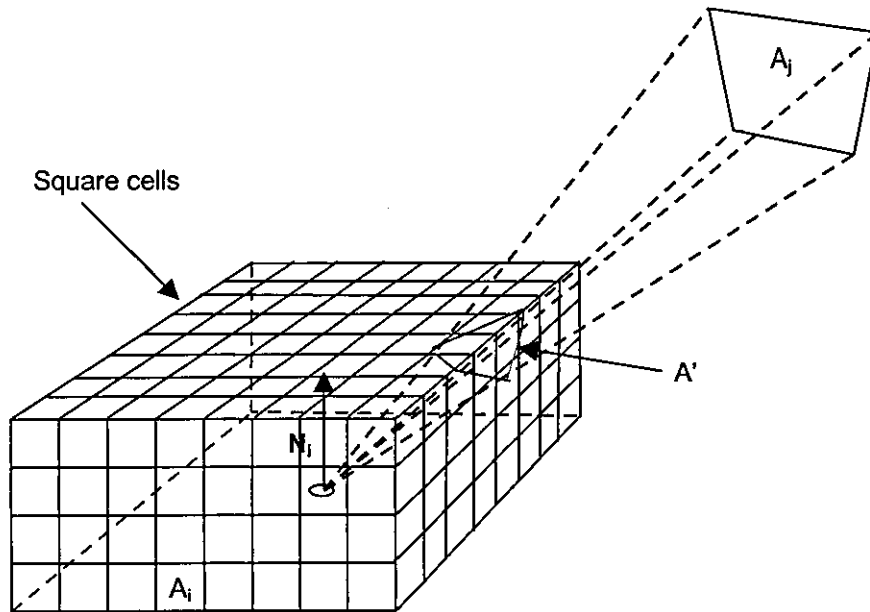
Taking an area average:

$$\begin{aligned} F_{A_i A_j} &= (1/A_i) \int_{A_i} F_{\Delta A_i A_j} dA_i \\ &= (1/A_i) \int_{A_i} \int_{A_j} [(\cos \phi_i \cos \phi_j) / \pi r^2] dA_j dA_i \quad \dots \text{eqn 9.5} \end{aligned}$$

Cohen and Greenberg also introduced a hemicube approach for approximating form factors and solving the hidden surface problem. In this technique, a cube is centred about  $dA_i$  with its top plane parallel to surface element  $dA_i$ . The patch  $A_j$ , whose form factor is to be determined, is projected onto the hemicube. The surface of the hemicube is subdivided into  $n \times n$  square cells. It has a unit height and a coordinate system with the origin at  $dA_i$ . (See figure 9.5.) The incremental contribution of each shaded pixel of area  $\Delta A$  and coordinates  $(x, y, z)$  to the total form factor is given by:

$$\Delta F_{A_i A_j} = [1 / (\pi (x^2 + y^2 + 1)^2)] \Delta A \quad \text{for top cells, and}$$

$$\Delta F_{A_i A_j} = [z / (\pi (x^2 + y^2 + 1)^2)] \Delta A \quad \text{for side cells.}$$



(Figure 9.5 – Hemicube)

If the projected patch,  $A'$ , covers  $m$  hemicube cells, the final form factor for patch  $A_j$  illuminating  $A_i$  is given by the sum of all cell contributions.

$$F_{A_i A_j} = \sum_{q=1}^m \Delta F_{A_i A_j} \quad \dots \text{eqn 9.6}$$

When more than one  $A_j$  patch shades a given cell, a depth sort determines which patch is closest. The incremental form factor contribution from the closest patch is used, thus resolving the hidden patch problem in a manner analogous to the z-buffer algorithm [CATM74].

The basic radiosity approach does suffer from a number of problems. For a start, it assumes only diffuse surfaces and therefore ignores specular reflections. Also the regular structure of the hemicube cells can produce aliasing effects. Because a scene with  $n$  patches generates an  $n \times n$  radiosity matrix, the problem size is of the order

$O(n^2)$ . For these reasons, a number of reformulations have been made to the radiosity rendering technique.

### 9.2.1 Progressive refinement

Cohen et al. [COHE88] developed a progressive refinement approach allowing form factors to be computed 'on-the-fly', hence providing a continually improving image that can be viewed whilst rendering. Processing starts with the brightest emitting patch, bounds it with a hemicube and shoots light through each hemicube cell to illuminate all  $n-1$  other patches. The brightest source is the patch with the largest  $B_i A_i$  product corresponding to the greatest energy emission. The radiosities are calculated for all patches due to illumination from the brightest patch, then the brightest resulting patch is found. Light is then shot from this patch to determine its influence on all other patches. The process continues until the image converges to a stable illumination balance.

Whereas the fundamental radiosity model considers the contribution of patch  $j$  to the radiosity of patch  $i$ :

$$B_i = \rho_i B_j F_{ij}$$

progressive refinement inverts this process by viewing patch  $i$  as the source which shoots light towards patch  $j$ :

$$B_j = \rho_j B_i F_{ji}$$

As:

$$F_{ij} A_i = F_{ji} A_j$$

the radiosity  $B_j$  due to  $B_i$  can be expressed as:

$$B_j = \rho_j B_i F_{ij} (A_i / A_j)$$

... eqn 9.7

The progressive refinement algorithm is:

```

repeat
  {select patch i with greatest stored energy}
  for j = 1 to n do
  begin
    {compute Fij relating i to each j patch}
    dBij = pj dBi Fij Ai / Aj
    dBj = dBj + dBij
    Bj = Bj + dBij
  end
  dBi = 0
until {image is good enough}

```

This algorithm has the advantage that the whole scene is updated on each iteration so it is possible to display the solution at each step. This gives the user control over the simulation and allows for interruption when the quality reaches a satisfactory level.

Amongst the other useful proposals, is that of Wallace et al. [WALL89] who reformulated form factor computation to be more convenient than the original hemicube approach. Their revision determines radiosity at vertices, which makes smooth polygon shading easier to perform.

The equation for the radiosity at *vertex 1* due to illumination by *source 2* is given by:

$$B_1 = \rho_1 B_2 A_2 (1/n) \sum_{i=1}^n [\delta_i (\cos\theta_{1i} \cos\theta_{2i}) / (\pi r_i^2 + A_2/n)] \quad \dots \text{eqn 9.8}$$

where  $n$  is the number of sample points on the source,

$\delta_i = 1$  if sample point is visible to vertex and 0 if occluded,

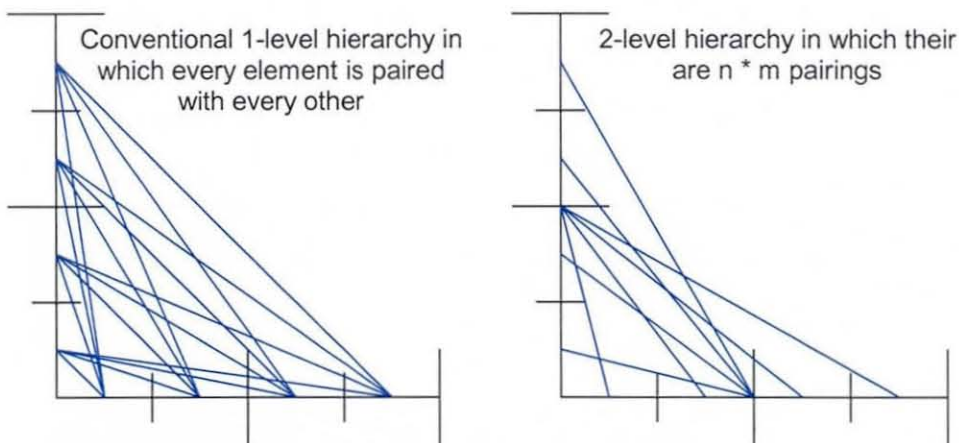
$r_i$  is the distance between the surfaces, and

$\cos\theta_{1i}$  and  $\cos\theta_{2i}$  are the angle of each surface with its normal.



### 9.2.2 Substructuring and adaptive subdivision

The quality of the solution achieved using radiosity is to a large extent dependent on the granularity of patch subdivision applied to the surfaces. Cohen et al. [COHE86] employed a two-level hierarchical technique called substructuring that not only subdivided the original surface polygons into a collection of patches but also subdivided each patch into a set of elements. This approach is based on the realisation that standard patch-to-patch calculations are often more accurate than necessary. Small-sized patches are only required for receiving patches where higher resolution is needed to capture the local details of the illumination. When a patch is being considered as the illuminator, higher resolution will not significantly alter its overall effect on the illumination of a distant patch. Hence Cohen's algorithm uses a two-level hierarchy to distinguish between the higher resolution needed when receiving energy, and the lower resolution when emitting or reflecting energy. The algorithm first subdivides the scene into a mesh of patches. Each patch is then further refined into smaller elements. When a patch is about to receive energy, the finer resolution is used to increase the accuracy of the radiosity representation. Patches, rather than the subdividing elements, are used when shooting energy thus reducing the number of iterations to be carried out. (See figure 9.6.)



(Figure 9.6 – Element pairings between perpendicular 2D patches)

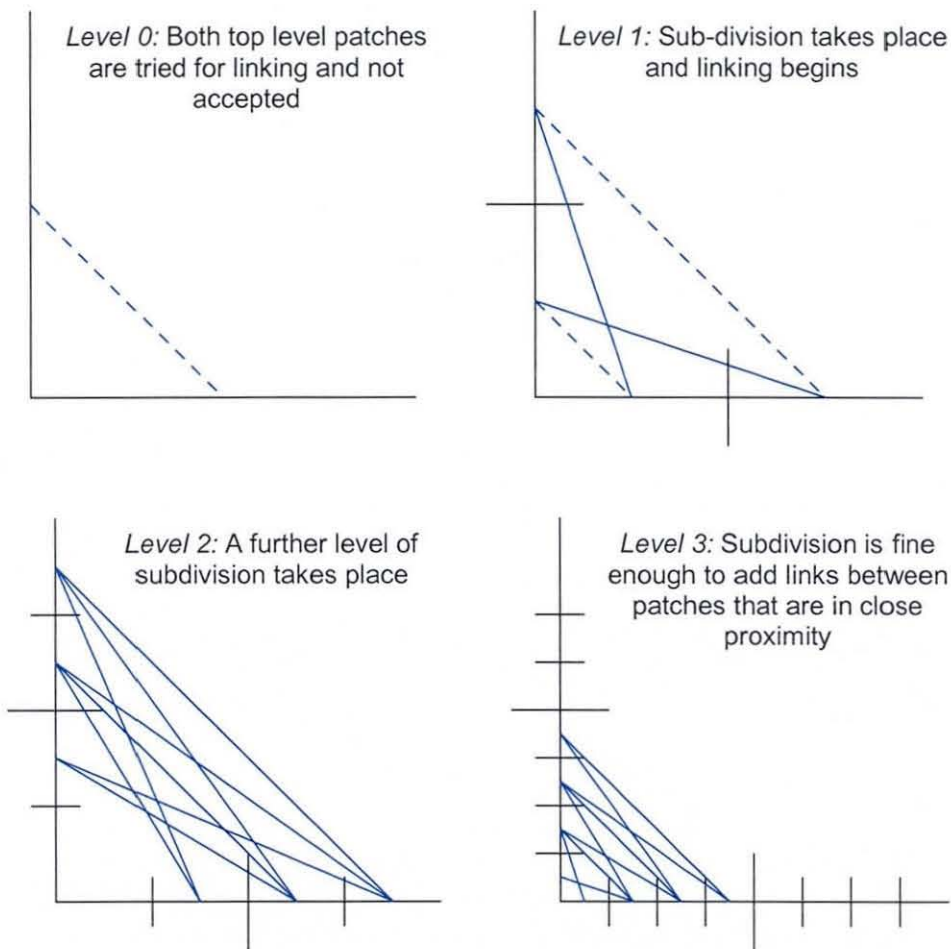
This technique does have the disadvantage that the mesh has to be created beforehand. An improvement is to subdivide each patch automatically as required in a process of adaptive mesh refinement. To achieve this a quadtree data structure is used to represent the subdivision of a patch into elements. This enables local neighbours to be easily accessed and the radiosity gradient over a patch can be used to determine subdivision granularity. While the gradient over a patch exceeds a given threshold, it can be subdivided and a new iteration started.

### 9.2.3 Hierarchical radiosity

The two-level hierarchical approach offers significant efficiency gains but it still over samples in areas where there is only slight variation in radiosity and under samples in areas of discontinuity such as at shadow boundaries.

Hanrahan et al. [HANR91] proposed a multi-level hierarchical scheme that subdivides each surface patch into a tree of sub-patches allowing each pairing to occur at an appropriate level. The approach was inspired by methods used to solve the *N-body* problem [BARN86]. The *N-body* problem relates to resolving the forces between  $N$  particles where each particle exerts a force on the others. Fast algorithms developed for this situation exploit the facts that the force between two particles only needs to be calculated to within a given precision and that the force due to a cluster of particles at some distance can be approximated without having to consider the influence of each individual particle. Hanrahan recognised the similarity between this situation and that of determining the influence of radiosities between patches. In both cases there are  $N(N-1)/2$  pairs of interactions and fall-off is proportional to  $1/r^2$ , where  $r$  represents the separation distance. However, whereas *N-body* algorithms begin with  $N$  particles and cluster them into larger and larger groups, global illumination determination needs to begin with a few large polygons and subdivide them, as appropriate, into smaller and smaller patches. Also, in the case of radiosity, occlusion needs to be taken into account.

With a hierarchical radiosity algorithm, there is no need to subdivide patches in advance. The starting point is a set of untessellated polygons that describe the scene. All polygons are compared with each other and subdivided as necessary. The polygons are subdivided according to a refinement criterion, which sets a threshold for the amount of energy transfer between them. The aim is to ensure that the energy transfer influence is approximately the same for all subdivision pairings. The result is a quadtree data structure that records the breakdown of a polygon into sub-elements and links between tree nodes reflect pairings. Links between leaf nodes represent the energy exchange between two single elements in the mesh. Links between nodes higher up the hierarchy represent interactions between averaged element clusters.



(Figure 9.7 – Pairings between perpendicular 2D patches using a multi-level hierarchy.

Solid lines show accepted links; dashed lines show attempts)



### 9.3 Previous common illumination approaches

One of the earliest attempts at addressing the common illumination problem is reported by Nakamae et al. [NAKA86]. When overlaying computer-generated images into a background photograph, they used the time and date of the photograph as well as the original scene longitude and latitude to determine the sun's position and hence calculate the shading of building and the shadows they would cast in the montage image. Clearly this simple approach takes no account of the possibility of multiple light sources, diffuse lighting or the complex illumination interactions that take place between real objects.

Fournier [FOUR94] tackled some of these issues using video images of a real scene. He modelled the principal objects in the scene with geometric primitive bounding shapes to produce an approximate 3D model. Surface elements were created from the sides of the bounding shapes and ray casting used to match projected screen pixels to surface elements. The assumption was made that pixel values in the scene images were proportional to the radiance at these points. A radiosity value was assigned to each visible element based on the average of all the pixels it contained. Element reflectivity was estimated based on weighting the average scene reflectivity in relation to the radiosity of neighbouring pixels on the principle that if the neighbourhood is darker, the reflectivity of the element in question is likely to be higher. Hidden surface elements were assigned the ambient radiosity and the average reflectivity. Light source positions and sizes were modelled as polygonal emitters. Light source emittance values were estimated by obtaining a global radiosity solution to provide a best fit for the radiosity estimates originally determined from the modelled scene surfaces. The sum of emittances was constrained to fit the estimate of total light intensity given by the ambient radiosity. The final model was used to calculate global illumination with respect to virtual objects that were added to the scene.

Fournier's method requires the user to specify the 3D shape of all objects in the scene. It uses projected pixel values to compute global illumination and therefore

needs a wide field of view. The user needs to specify the positions of all light sources.

Drettakis et al. [DRET98] extended Fournier's work. They applied vision-processing techniques using a calibration pattern to compute viewing parameters and to create a geometric model of the scene by point correspondences between 12 images. (See discussion of scene reconstruction strategies in chapter 8.) They also used fast hierarchical and incremental update radiosity techniques improving computational efficiency of global illumination. However, the basic principles employed to estimate radiosities was exactly the same as Fournier's and hence did not allow for any unknown surfaces outside the field of view. Light source positions must be known before any global illumination model can be determined.

Debevec [DEBE98] presented a method in which the real scene is partitioned into distant and local components. The local scene represents the objects in the vicinity of where the synthetic object is to be placed, while the distant scene consists of the surrounding environment. On the assumption that no part of the distant scene is affected by light reflecting or emitting from the local scene, a light-based radiance model of the distant scene is acquired using a light probe. The probe is a polished steel ball placed in the scene near to the proposed location of the synthetic object(s). The ball is photographed at different exposures to obtain a full dynamic range radiance map of the surroundings (figure 9.8). The geometry and material properties of objects in the local scene must be modelled. However, for the distant scene, the acquired radiance information is mapped onto a box to represent the inside walls and ceiling of the room. This coupled with the information about the local scene and the synthetic object(s) is used to render the composite scene.





(Figure 9.8 – High dynamic range radiance maps produced by photographing probe at different exposure settings)

A similar approach to Debevec's uses a pair of CCD cameras fitted with fish-eye lenses and positioned at the proposed virtual object location [SATO99]. Images from the cameras are used to acquire a radiance distribution map of the environment and, as two cameras are used in a stereo pairing, some geometric information about the surroundings can be deduced.

Unfortunately the early approaches to solving the common illumination problem rely on knowledge of the size and location of light sources. Recent methods address this issue but require the environment radiance map to be obtained by placing a probe or cameras into the real scene prior to compositing with synthetic objects.

The next sections describe a method for automatically acquiring a plausible global illumination model based only on the information available within the local scene.

#### **9.4 Finding a plausible common illumination solution**

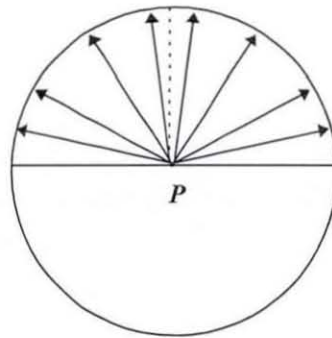
In this section the theoretical basis for a novel approach to solving the common illumination problem is presented. The method seeks to infer a plausible global illumination solution from a real scene assuming only that its visible geometry and surface material properties are known.

For some scenes, where the object geometry and surface properties are known or can be determined, it is possible, in principle, to track backwards along imaginary rays from the location of highlights and cast shadows, to provide an indication of the directions to the original light sources. However, such ray-casting techniques cannot deal effectively with light sources that have finite area and, in reality, few if any light sources fulfil the qualification of being point sources. Indeed, it is almost always the case that our environment is illuminated with light that is diffusely reflected from surface to surface. Thus, so-called photo-realistic computer graphic renderers will usually employ ray tracing to achieve specular illumination effects, in combination with a diffuse radiation modelling technique based on a radiosity model.

While, in principle, it is quite straightforward to reverse ray trace, inferring an illumination model from the diffuse illumination is much more difficult to realise. In fact, it is impossible to infer original light source size, distance, orientation and brightness with absolute certainty based on diffuse reflections, since diffuse surfaces scatter incident light in a multiplicity of directions and therefore appearance cannot be used to deduce the exact parameters of illuminating rays. Having said that a perfect solution is impossible, this is not to imply that we cannot infer some plausible approximation.

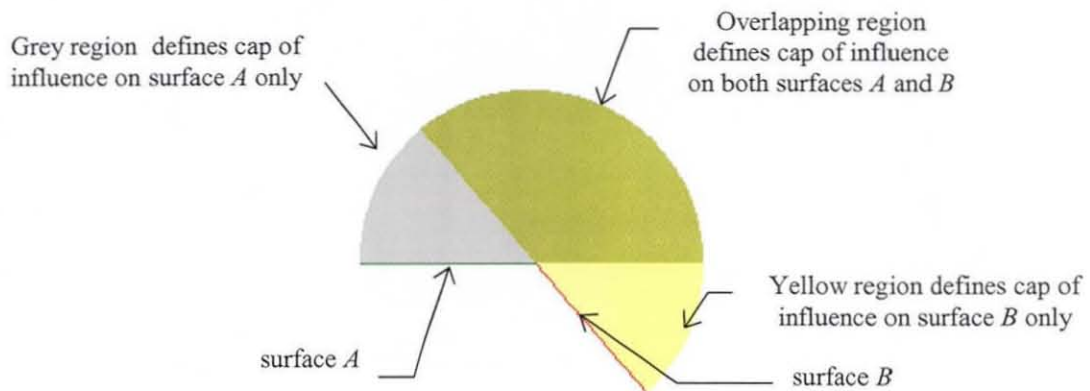
##### **9.4.1 Theoretical basis**

In the case of diffuse reflection at a point we can, at best, say that there must be a light source or sources to the illuminated side of the surface. In effect this narrows the attributable source direction down to a hemisphere centred on the surface point being considered.



(Figure 9.9 – Range of possible directions of light sources causing diffuse reflection at point  $P$  on a surface)

As illustrated in figure 9.9, given the intensity at  $P$ , all we can conclude is that this must be due to sources of light somewhere in the direction of the upper hemisphere surface. Extending this to two points, each on a separate surface,  $A$  and  $B$ , we could imagine a sphere constructed to enclose the two surfaces. (See figure 9.10.)



(Figure 9.10 – Range of possible directions of light sources causing diffuse reflection at point  $P$  on a surface)

Any diffuse light reaching surface  $A$  must have come from the direction of the spherical cap surrounding its illuminated side, shown in the diagram as enclosing the grey hemispherical region. Similarly, any diffuse light reaching surface  $B$  must originate on the hemispherical surface bounding the yellow region. If, for example, surface  $A$  is in darkness, we can infer that the diffuse light radiation arriving at



surface  $B$  must have come from the direction of  $B$ 's cap minus  $A$ 's cap, i.e. the region of the enclosing sphere's surface that bounds the pale yellow region in the above diagram. Thus it appears feasible, given sufficient information from visible surfaces, to narrow down the possible distribution of lighting around the inside of the enclosing sphere in a way that can explain the perceived illumination of the enclosed objects. Any solution that can account for the perceived illumination could, in theory, then be used as a common illumination model that should produce plausible results, even though it may not truly match the original light source disposition.

In a situation where we have a view of only a portion of the real world, some object surfaces will be visible whilst others, occluded by other entities or facing away from the viewer, will not be seen. In relation to the diffuse lighting within a scene, all these surfaces (assuming they are to some degree reflective) will have an effect. Other surfaces, outside the field of vision, are also almost certain to be present and, hence, will contribute to the global balance of diffuse radiation.

The nature and geometry of surfaces within the field of view may be known or be deduced but surfaces outside may be unknown and substantially unknowable. This prevents determination of a satisfactory illumination model for the viewed scene. Nevertheless, under certain conditions it should be possible to infer enough about the global illumination from the intensity across visible surfaces and hence to re-illuminate the scene convincingly after the addition of virtual objects or other changes in inter-object geometry. This may be accomplished by enclosing the visible extent of the real world in an imaginary containment sphere that, as a first step, collects all radiation that cannot be accounted for by the known radiosity of surface patches within the scene. After the introduction of virtual objects or other changes within the scene, individual areas on the inside surface of the containment sphere can be treated as emitters to re-illuminate the scene in a conventional radiosity process.

To model this situation, the radiosity equation 9.3 needs to be revised to account for unknown surface patches. This is done by splitting the summation term into three parts:

$$B_i = E_i + \rho_i \left[ \sum_{j=1}^r B_j F_{ij} + \sum_{k=(1+r)}^{(r+s)} B_k F_{ik} + \sum_{m=(1+r+s)}^{(r+s+t)} B_m F_{im} \right] \quad \dots \text{eqn 9.9}$$

where;

- $r$  represents the number of visible surfaces and therefore of known radiosity;
- $s$  represents the number of invisible surfaces but which are known to be within the scene; and
- $t$  represents the number of facets within the containing enclosure. (These facets are assumed to have zero reflectivity, but radiosity emittance,  $B_m$ .)

The sum of the enclosure facet emittances must account for all light radiation that cannot be explained by the surfaces known to be within the scene.

The first summation,  $\sum B_j F_{ij}$ , represents the contribution to the radiosity of patch  $i$  made by all visible patches within the extent of the scene. The second summation,  $\sum B_k F_{ik}$ , represents the contribution to the radiosity of patch  $i$  made by all invisible patches within the extent of the scene. The third,  $\sum B_m F_{im}$ , represents the contribution to the radiosity of patch  $i$  made by all patches making up the containment sphere. Together these must account for all the visible radiosity.

Evaluation of the first summation is possible by measurement of the intensity of each visible patch. However, the other two summations contain unknown radiosity values.

Since the enclosed world contains  $(r+s)$  surface patches, producing an equation for each of these patches gives  $(r+s)$  equations, which will contain  $(s+t)$  unknowns.



However, if we ensure, by appropriate patch division of the containment sphere, that  $(r+s) \geq (s+t)$ , i.e.  $(r \geq t)$  then, in principle, the equations should be solvable.

Equation 9.9 can be expressed in matrix form:

$$\mathbf{A} \mathbf{x} = \mathbf{b}$$

as follows:

$$\begin{bmatrix} \rho_1 F_{1(r+1)} & \dots & \rho_1 F_{1(r+s)} & \rho_1 F_{1(r+s+1)} & \dots & \rho_1 F_{1(r+s+t)} \\ \dots & & & & & \\ \dots & & & & & \\ \rho_r F_{r(r+1)} & \dots & \rho_r F_{r(r+s)} & \rho_r F_{r(r+s+1)} & \dots & \rho_r F_{r(r+s+t)} \\ 1 - \rho_{(r+1)} F_{(r+1)(r+1)} & \dots & -\rho_{(r+1)} F_{(r+1)(r+s)} & -\rho_{(r+1)} F_{(r+1)(r+s+1)} & \dots & -\rho_{(r+1)} F_{(r+1)(r+s+t)} \\ \dots & & & & & \\ \dots & & & & & \\ -\rho_{(r+s)} F_{(r+s)(r+1)} & \dots & 1 - \rho_{(r+s)} F_{(r+s)(r+s)} & -\rho_{(r+s)} F_{(r+s)(r+s+1)} & \dots & -\rho_{(r+s)} F_{(r+s)(r+s+t)} \end{bmatrix} \begin{bmatrix} B_{(r+1)} \\ \dots \\ \dots \\ B_{(r+s)} \\ B_{(r+s+1)} \\ \dots \\ \dots \\ B_{(r+s+t)} \end{bmatrix}$$

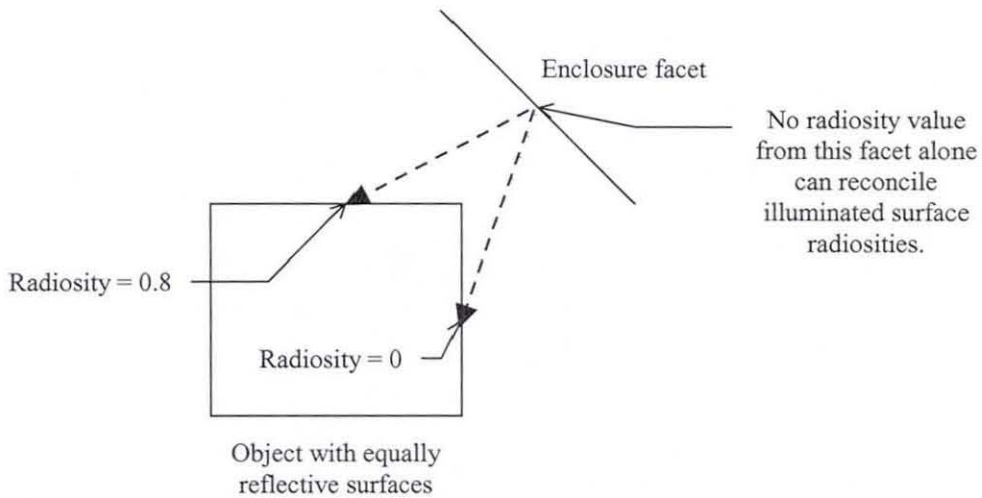
$$= \begin{bmatrix} (B_1 - E_1) - \rho_1 [B_1 F_{11} & \dots & -B_r F_{1r}] \\ \dots & & \\ \dots & & \\ (B_r - E_r) - \rho_r [B_1 F_{r1} & \dots & -B_r F_{rr}] \\ E_{(r+1)} + \rho_{(r+1)} [B_1 F_{(r+1)1} & \dots & + B_r F_{(r+1)r}] \\ \dots & & \\ \dots & & \\ E_{(r+s)} + \rho_{(r+s)} [B_1 F_{(r+s)1} & \dots & + B_r F_{(r+s)r}] \end{bmatrix}$$

... eqn 9.10

The solution for vector  $\mathbf{x}$  is given by:

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b} \quad (\text{where } \mathbf{A}^{-1} \text{ is the inverse of } \mathbf{A}.)$$

Unfortunately, things are not so straightforward as they were for standard radiosity. The system of linear equations represented by equation 9.10 are likely to be inconsistent, having no solution that will satisfy them exactly, i.e.  $\mathbf{A}$  has no inverse. This is because the form factors relating to facets of the containing sphere will almost certainly not match those of the original light source(s). The sphere's facets will not, except by some extremely remote chance, match the area, orientation and distance of the original sources. For example, one of the enclosing sphere's facets could, according to its form factors, have equal influence over two world patches; one of which is in darkness and the other well lit as illustrated in figure 9.11.



(Figure 9.11 – Influence of a single facet on two patches may be impossible to reconcile)

No single facet radiosity can reconcile the situation; the corresponding linear equations are inconsistent.

As the equations are likely to be over-constrained and no single 'perfect' solution will be obtainable, we must content ourselves with finding a 'best-fit' solution and can relax the principle that the number of equations and unknowns must match. We also need to avoid the possibility of enclosure facets being assigned negative radiosities. If all we wanted to do was re-illuminate the same scene unaltered, this may well give appropriate results mathematically, but insertion of virtual objects

could create anomalies. For example, if a new object obscures an original surface from its negatively contributing facet, the original surface will become brighter and the new object could have negative intensity, the latter being a physical impossibility in terms of the real world.

#### 9.4.2 Solving the equations

The linear system described by equation 9.10 presents two problems:

1. the equations are likely to be inconsistent; having no feasible solution, and
2. a 'best-fit' solution is likely to result in negative radiosities, which is a physical impossibility.

These problems and their resolution can be illustrated using simple examples.

Firstly, assume we have two linear equations to satisfy:

$$\begin{aligned}x &= 2 \\x &= 4\end{aligned}$$

Obviously, these equations are inconsistent with one another and have no feasible solution for  $x$ . However, we could find a 'best-fit' solution, minimising the error in both equations by assigning  $x$  the value 3. With more equations and variables, this averaging process becomes a bit more complex but can be achieved by first introducing artificial variables to make the equations feasible, and then by minimising the values of these variables.

In the above example, introducing artificial variables,  $e_1$  and  $e_2$ , gives:

$$\begin{aligned}x + e_1 &= 2 \\x + e_2 &= 4\end{aligned}$$

and in  $\mathbf{Ax} = \mathbf{b}$  matrix form:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

The equations now have at least one solution; that is, when  $x = 0$ ,  $e_1 = 2$  and  $e_2 = 4$ .

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$|\mathbf{e}|^2$  gives a measure of the error which, in this case, is:

$$(e_1^2 + e_2^2) = (4 + 16) = 20$$

However, when  $x = 3$ ,  $e_1 = -1$  and  $e_2 = 1$ :

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

the error is minimised at:

$$|\mathbf{e}|^2 = (1 + 1) = 2$$

The second problem, that of negative values, can be illustrated with another example based on the two plane equations:

$$\begin{aligned} 2x + y + z &= 1 \\ x + 3y + 2z &= 7 \end{aligned}$$

... eqns 9.11

which rewritten in matrix form,  $\mathbf{Ax} = \mathbf{b}$ , gives:

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

We have fewer equations than unknowns and there are many possible solutions. In such situations we can determine a least-squares best fit using Singular Value Decomposition to obtain a pseudo-inverse, denoted by  $\mathbf{A}^+$  [STRA88] such that:

$$\mathbf{x}^+ = \mathbf{A}^+ \mathbf{b} \quad (\text{where } \mathbf{x}^+ \text{ represents the least-squares solution})$$

In the above example, Singular Value Decomposition of  $\mathbf{A}$  gives:

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} -0.5019 & -0.8649 \\ -0.8649 & 0.5019 \end{bmatrix} * \begin{bmatrix} 4.250 & 0 & 0 \\ 0 & 1.392 & 0 \end{bmatrix} * \begin{bmatrix} -0.4397 & -0.7286 & -0.5251 \\ -0.8821 & 0.4604 & 0.0998 \\ -0.1690 & -0.5071 & 0.8452 \end{bmatrix}$$

$\mathbf{u} \qquad \qquad \qquad \Sigma \qquad \qquad \qquad \mathbf{v}^T$



The pseudo-inverse,  $\mathbf{A}^+$ , is given by  $\mathbf{v} \times \Sigma^{-1} \times \mathbf{u}^T$  thus giving:

$$\begin{bmatrix} -0.4397 & -0.8821 & -0.1690 \\ -0.7286 & 0.4604 & -0.5071 \\ -0.5251 & 0.0998 & 0.8452 \end{bmatrix} \times \begin{bmatrix} 0.2353 & 0 \\ 0 & 0.7183 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} -0.5019 & -0.8649 \\ -0.8649 & 0.5019 \end{bmatrix}$$

$\mathbf{v} \qquad \qquad \qquad \Sigma^{-1} \qquad \qquad \qquad \mathbf{u}^T$

$$= \begin{bmatrix} 0.6 & -0.229 \\ -0.2 & 0.314 \\ 0 & 0.143 \end{bmatrix}$$

The least-squares solution is given by:

$$\mathbf{x}^+ = \mathbf{A}^+ \mathbf{b}$$

thus:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Assuming we wish to constrain the solutions such that:

$$x \geq 0, y \geq 0, \text{ and } z \geq 0$$

the problem of finding a least-squares solution becomes non-linear.

In general terms, we now wish to minimise the error:

$$|\mathbf{Ax} - \mathbf{b}|^2$$

subject to the constraints:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} && \dots \text{ eqn 9.12} \\ \mathbf{x} &\geq 0 \quad (\text{i.e. } x_1, x_2, x_3, \dots \geq 0) \end{aligned}$$

Such problems, where the objective function (the function to be minimised) is quadratic and the constraints are linear are known as Quadratic Programming problems and can be solved. (Texts on non-linear optimization discuss such problems and their resolution [eg WISM78].)

Utilising a constrained, non-linear, optimization algorithm developed by Lawrence et al. [LAWR97] of the University of Maryland, to resolve equations 9.11 gives a solution:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 2.2 \\ 0 \end{bmatrix}$$

which amounts to a least-squares error:

$$|\mathbf{Ax} - \mathbf{b}|^2 = (2.2 - 1)^2 + (6.6 - 7)^2 = 1.6$$

For the reverse radiosity situation, to contend with inconsistency, an artificial variable can be added to each of the equality constraints in equation 9.12. This ensures that there is a feasible solution for each constraint, i.e. if all other variables,  $\mathbf{x}$ , are set to zero, the vector of artificial variables,  $\mathbf{e}$ , becomes equal to vector  $\mathbf{b}$ , since:

$$\mathbf{Ax} + \mathbf{e} = \mathbf{b}$$

Since any value taken by the artificial variables constitutes a degree of error, the objective function to minimise the error can be re-expressed as:

$$|e|^2 \quad \text{where } |e|^2 = |\mathbf{Ax} - \mathbf{b}|^2$$

So, the Quadratic Programming problem can be expressed as:

$$\text{Minimise the objective function: } |e|^2$$

Subject to the constraints:

$$\begin{aligned} \mathbf{Ax} + \mathbf{e} &= \mathbf{b} && \dots \text{ eqn 9.13} \\ \mathbf{x} &\geq 0 \end{aligned}$$

where  $\mathbf{Ax} = \mathbf{b}$  is fully described by equation 9.10.

### 9.4.3 Solving Quadratic Programming problems

Quadratic Programming problems are difficult to solve. However, algorithms exist and some implementations are freely available. FSQP Version 2.5, written in C and developed by the Institute for Systems Research (ISR) at the University of Maryland [LAWR97] was obtained from AEM Design [AEM03] and used in the tests described in section 9.6. It is an implementation of two algorithms based on Sequential Quadratic Programming (SQP), modified to generate feasible iterates. Processing is in two phases; the first generates an iterate satisfying all linear constraints and non-linear inequality constraints, and the second seeks to minimize the maximum of the objectives, while maintaining satisfaction of the constraints.

## 9.5 Viewing and rendering software implementation

To provide a test environment with fully accessible source code, a 3D viewing and rendering system was developed. This implements a class called **Port3D**, an instance of which defines a 3D port that can be drawn into using real-world 3D coordinates. Progressive refinement radiosity is used for rendering. The implementation is provided on the CD-ROM attached as appendix F.

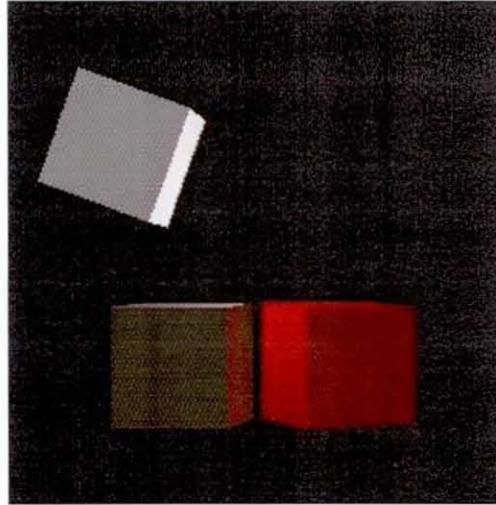
Scenes are described as text files consisting of composite and primitive objects. Composite objects can be made from other composites or from primitives. The Scene is treated as a Composite object and the file format is:

```
[No of objects]
{for each object}
    [primitive | composite]
    [object name]
    [xScale] [yScale] [zScale]
    [xRotate] [yRotate] [zRotate]
    [xTranslate] [yTranslate] [zTranslate]
{end for}
```

Primitive objects have the following format:

```
[No of vertices]
{for each vertex}
    [xCoord] [yCoord] [zCoord]
{end for each vertex}
[No of surfaces]
{for each surface}
    [surface id]
    [red reflectivity] [green reflectivity]
                                [blue reflectivity]
    [red emittance] [green emittance] [blue emittance]
    [No of polygons in surface]
    [polygon id] [counter-clockwise list of vertex ids]
{end for each surface}
```

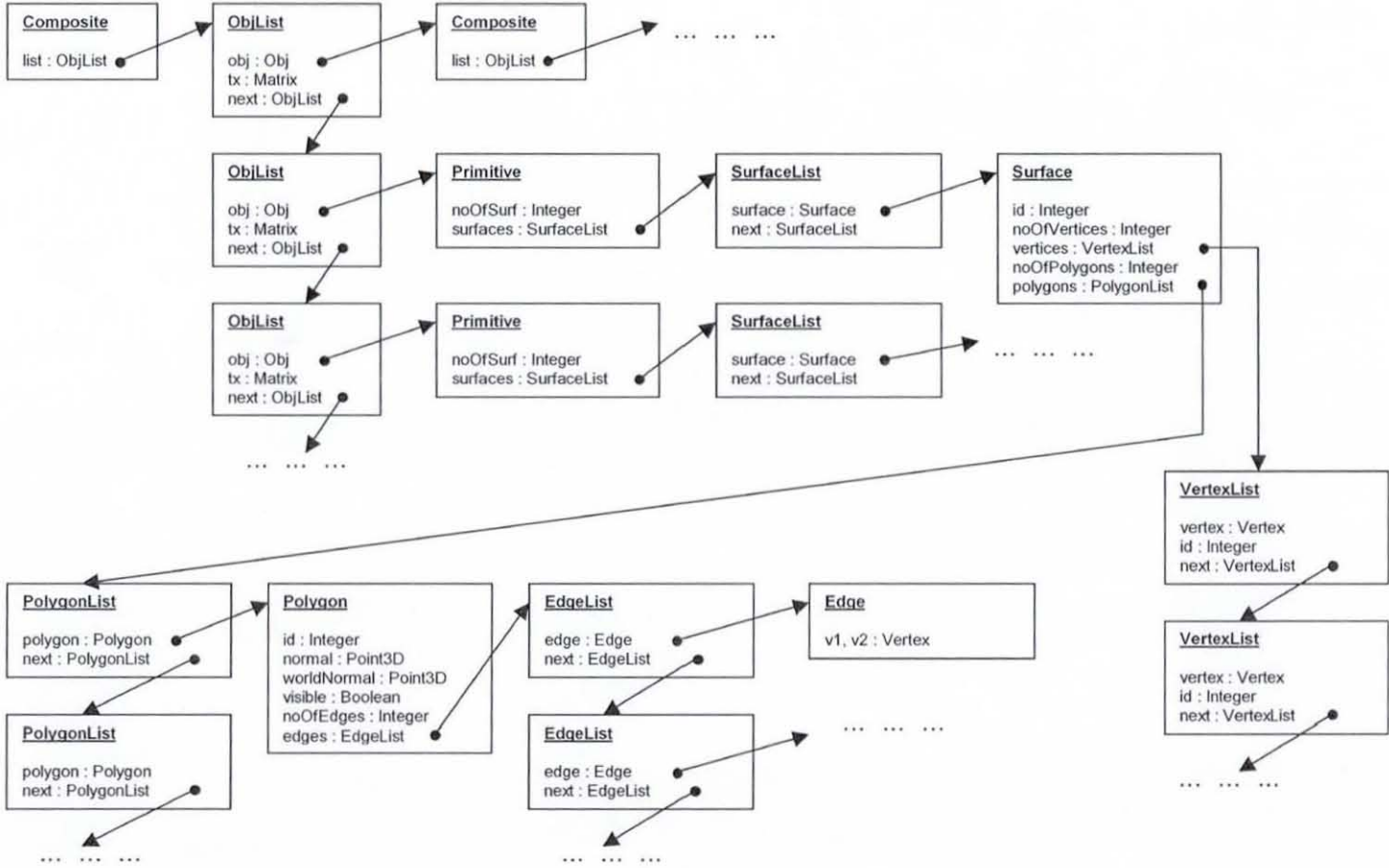
Figure 9.12 shows the rendering of a red and a white cube is close proximity illustrating the colour bleeding effect.



*(Figure 9.12 – Radiosity rendering showing colour bleeding)*

As a scene file is read, a scene data structure as shown in figure 9.13 is built. This allows transformations applied to composite objects to ripple through to transform all its constituent primitives.





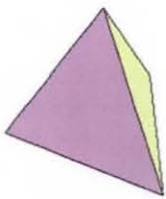
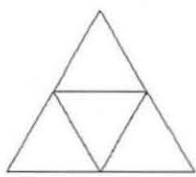
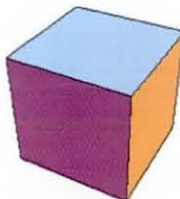
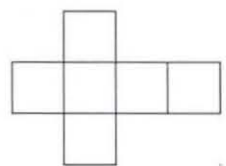
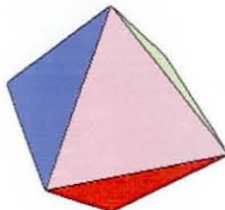
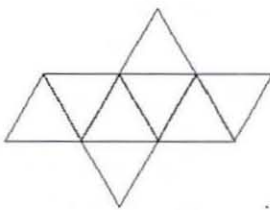
(Figure 9.13 – Scene data structure used by rendering software)

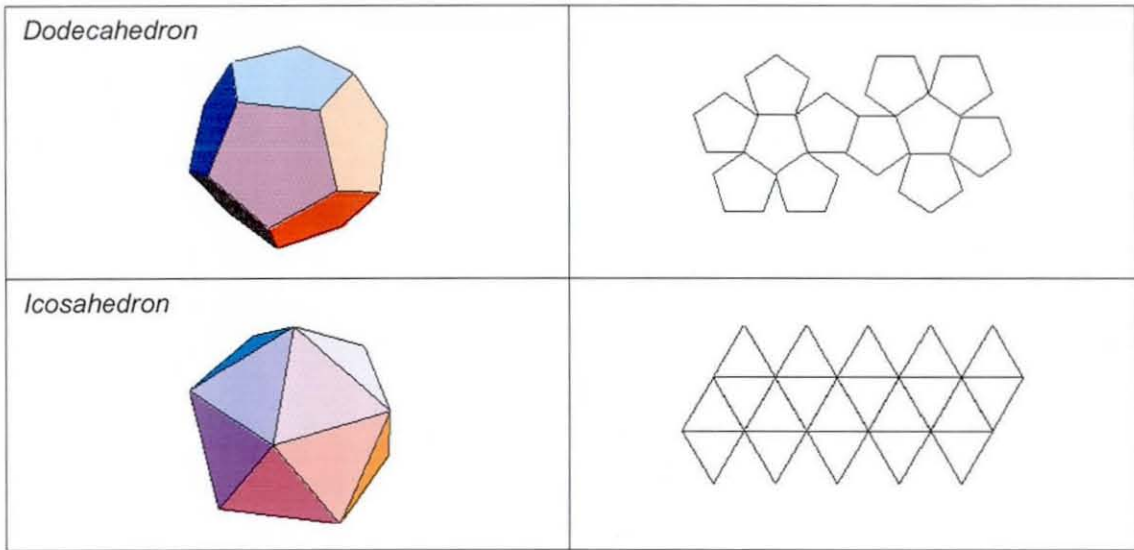
## 9.6 Testing

This section describes empirical verification of the common illumination inference theory using a simple synthetic scene and a real scene.

### 9.6.1 Constructing the enclosure

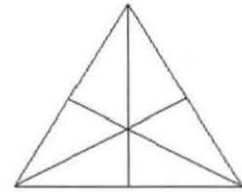
How to best construct a suitable enclosure is, in itself, not a straightforward issue. We could start by forming some approximation to a surrounding sphere with an appropriate number of polygonal facets. If a particular number of equal facets are required this becomes difficult, as there is no known way to divide a sphere's surface into any chosen number of regular facets. However, it is possible to achieve successively higher resolution approximations starting from any of the five congruent, regular, convex polyhedra; the Platonic Solids. (See figure 9.14.)

|   |  |
|---|--|
| <p><i>Tetrahedron</i></p>  |  |
| <p><i>Cube</i></p>         |  |
| <p><i>Octahedron</i></p>   |  |



(Figure 9.14 – Platonic solids )

If, for example, we start with an octahedron, which has eight triangular faces, we could bisect each angle and pull each new vertex out to the full radius to form six new triangular facets from each original face.



Alternatively, we could bisect each edge, and pull all new vertices out to the full radius, turning each original triangular face into four new triangles. This process could be repeated any desired number of times for each of the faces giving a series (8, 32, 128, ..... ) of polyhedra with  $2 \times (4)^n$  faces. A similar process of subdivision, based on an icosahedron, would produce a series (20, 80, 320, ..... ) with  $5 \times (4)^n$  faces.

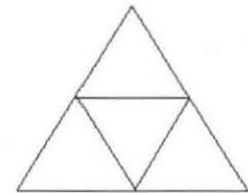


Figure 9.15 –  
Alternative sub-  
division of  
triangular facets

Ultimately, it may be possible to provide improved pixellation of the enclosure by producing irregular polyhedra that increase the number of facets corresponding to areas of most interest. At its crudest, this could entail simply slicing away planes through the enclosure that correspond to surfaces that are known to bound the real environment, such as walls and floors. Additionally, it should be possible to

maximise the effectiveness of enclosure patch size and position by an appropriate distribution of containment sphere patch division based on the prevailing disposition of scene patches.

### ***9.6.2 The illumination inference algorithm***

The overall algorithm proposed here, starts with a restricted view of a scene for which an *a priori* geometric and surface property model has been acquired, and takes the following steps.

#### *Step 1*

Divide the known surfaces into a number of patches.

#### *Step 2*

For each patch, measure its intensity and hence determine the radiation incident upon it.

#### *Step 3*

Using a standard radiosity model, determine the light radiation incident on each patch that can be accounted for by known patches, and subtract this radiosity from that of each patch. (The remaining radiosity must therefore be due to radiated energy from outside the restricted view of the scene.)

#### *Step 4*

Completely enclose the scene within a faceted enclosure.

#### *Step 5*

Determine the enclosure facet radiosities necessary to account for the remaining radiosity within the scene using Quadratic Programming techniques.

#### *Step 6*

Alter the geometry of the original scene or introduce new objects and re-illuminate using the enclosure.

### 9.6.3 A synthesised example

In this section, the novel steps, 4 to 6, of the above algorithm are demonstrated with a synthesised example scene. The starting point is a plane surface object,  $S$ , defined by its corner vertices at 3D world coordinates:

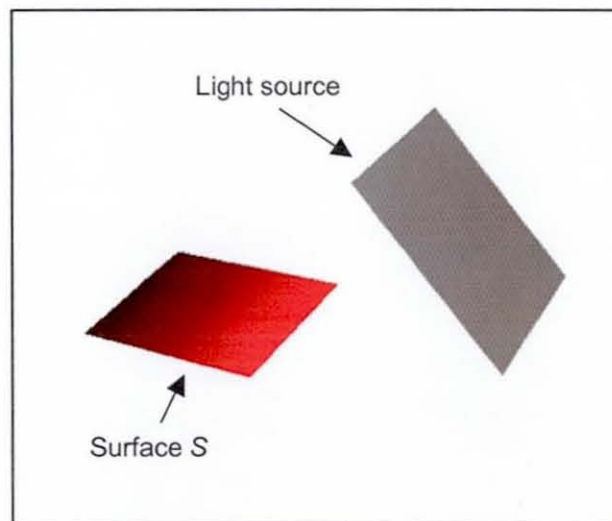
$$(3,4,1), (5,4,1), (3,4,-1) \text{ and } (5,4,-1)$$

The surface is divided into 16 square patches each of which has an area of 0.25 units.

Surface  $S$  is illuminated by a plane white light source with boundary vertices at coordinates:

$$(6,6,-1), (7.414, 4.5873, -1), (6,6,1) \text{ and } (7.414, 4.5873, 1)$$

These coordinates put the light source at  $45^\circ$  to the plane of  $S$  as shown in figure 9.16. The light source is considered to comprise of four equal sized facets each of area 1 unit.

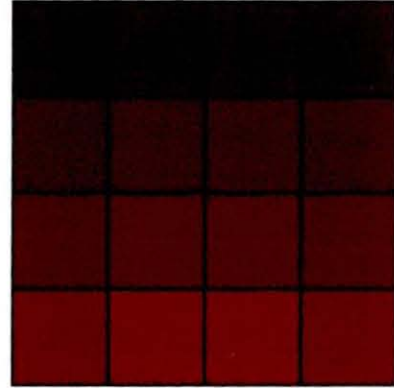


(Figure 9.16 – Light source illuminating Surface S)



Applying conventional radiosity techniques to calculate the intensity at the centre of each surface patch due to the light source gives the following relative intensity values. To the right of the table in figure 9.17, these have been converted into actual intensity values, assuming surface S to be red.

|        |        |        |        |
|--------|--------|--------|--------|
| 0.0276 | 0.0295 | 0.0295 | 0.0276 |
| 0.0403 | 0.0437 | 0.0437 | 0.0403 |
| 0.0604 | 0.0668 | 0.0668 | 0.0604 |
| 0.0918 | 0.1041 | 0.1041 | 0.0918 |



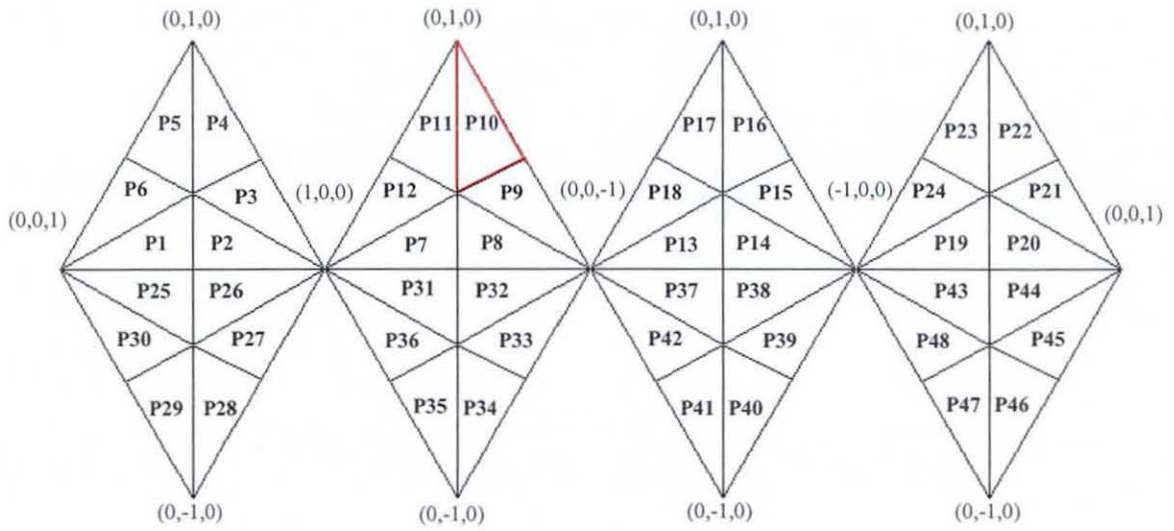
(Figure 9.17 – Surface S, red radiosities and appearance)

This forms the starting point of visible, known radiosities and, from this juncture, we assume that the original light source is unknown, being outside the field of view.

### ***Creating the enclosure***

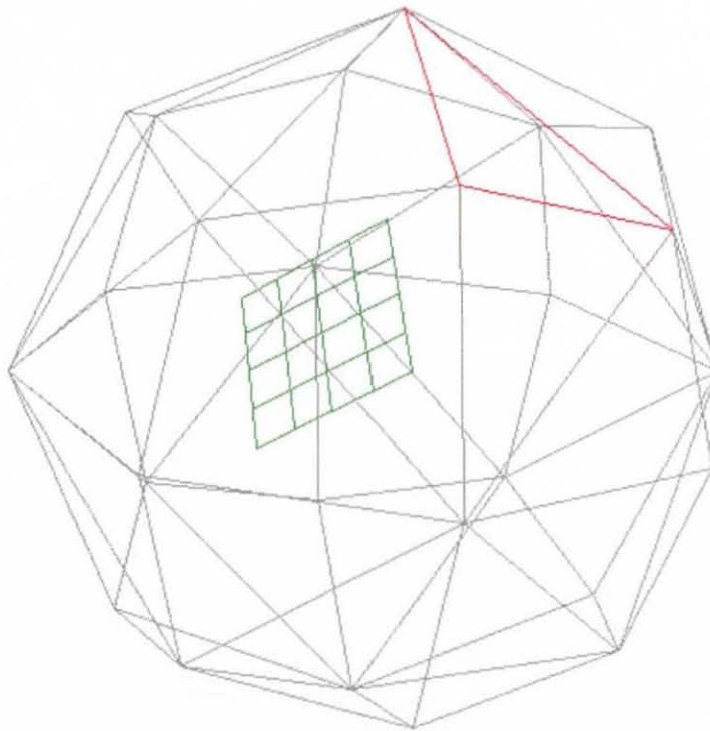
An enclosure for surface S was constructed based on subdivision of each face of an octahedron into six triangles, giving a 48-facet enclosure. For convenience, the enclosure was given a radius of 1 unit and the enclosed surface S was translated, scaled and rotated to fit within.

Figure 9.18 shows the enclosure facets laid flat as viewed from inside the enclosure. For identification purposes, the facets are labelled P1 to P48.



(Figure 9.18 – Faceted enclosure laid flat)

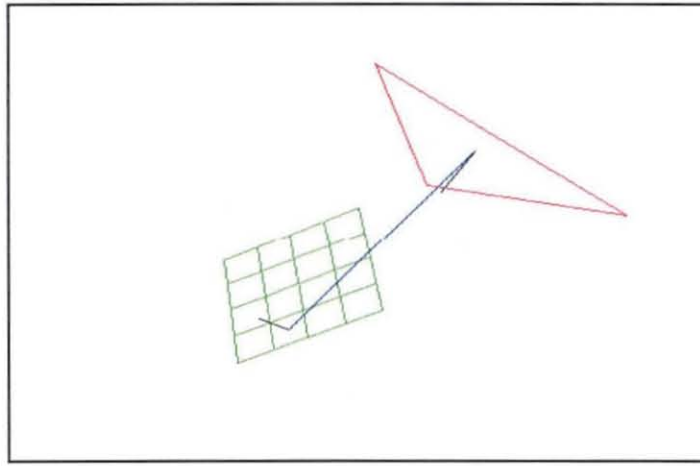
A three-dimensional wireframe view of the enclosure and original surface is reproduced as figure 9.19.



(Figure 9.19 – Surface S inside enclosure)

**Calculating the form factors**

One facet in figure 9.19 is outlined in red to show correspondence with figure 9.20, which shows the construction for form factor calculation between one enclosure facet and a single surface patch. The short black lines mark the normals. The respective areas, separation distance and orientation are used in the form factor calculation using Wallace's equation. (See equation 9.8.)



(Figure 9.20 – Construction for form factor calculation)

**Inferring the facet radiosities**

After calculating the form factors between each enclosure facet (P1 to P48) and each Surface S patch were calculated, and equation 9.10 applied with:

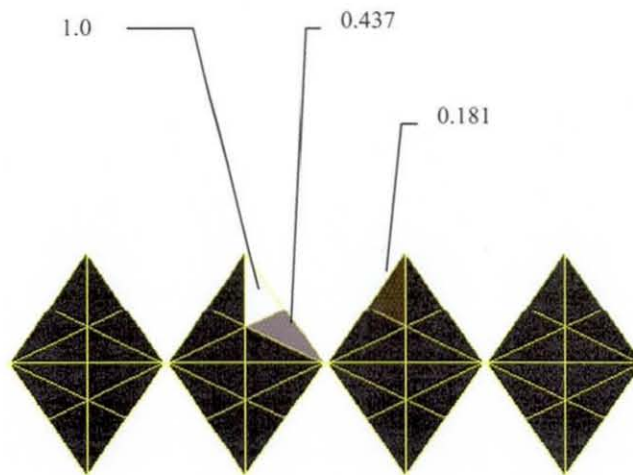
$r = 16$  (the number of visible surface patches - radiosity known),

$s = 0$  (the number of invisible surface patches - assuming S is single sided),

$t = 48$  (the enclosure facets - radiosity to be determined).

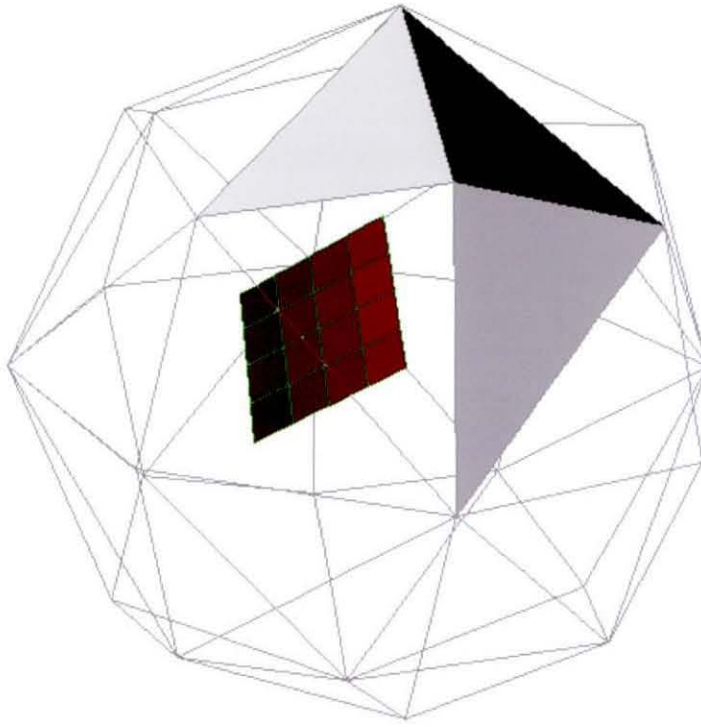
In this case, we would expect **A** to be a 16 x 48 matrix, **x** a 48 x 1 vector, and **b** a 16 x 1 vector. However, after introducing artificial variables and adding their associated columns to **A**, the result becomes a 16 x 64 matrix. Similarly, adding artificial variable rows to **x** results in a 64 x 1 vector.

Using a Quadratic Programming algorithm, implemented as FSQP Version 2.5 and described in section 9.4.3, to constrain the 48 enclosure patches to each lie between 0 and 1, whilst minimising the magnitude of the artificial variables, gives relative enclosure facet radiosities, as shown in figure 9.21. Radiosities of the unlabeled facets is 0.



(Figure 9.21 – Enclosure facet radiosities)

Figure 9.22 shows the surface and the enclosure. So they can be seen against the white paper background, the three enclosure facets determined to account for illumination of the surface are shaded in reverse intensity to their radiosity. Thus, the black shaded triangle corresponds to the brightest facet.

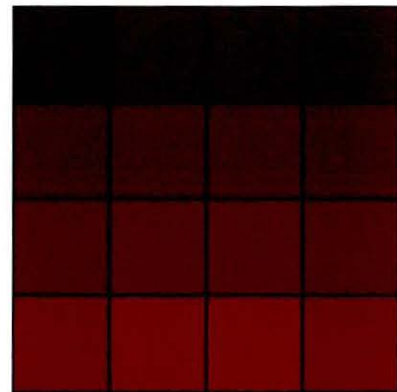


(Figure 9.22 – Surface *S* and enclosure showing lit facets)

### *Re-illuminating the surface*

Using normal radiosity techniques to re-illuminate the original surface *S*, this enclosure gives the facet radiosities shown in figure 9.23, which compares very closely to the original surface illumination in figure 9.17.

|        |        |        |        |
|--------|--------|--------|--------|
| 0.0290 | 0.0307 | 0.0306 | 0.0286 |
| 0.0416 | 0.0448 | 0.0447 | 0.0411 |
| 0.0612 | 0.0669 | 0.0670 | 0.0606 |
| 0.0915 | 0.1019 | 0.1026 | 0.0908 |

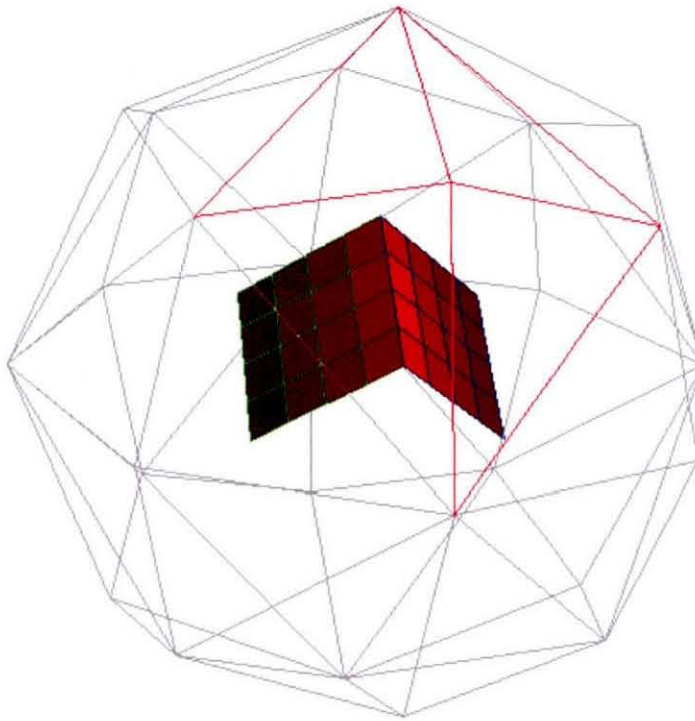


(Figure 9.23 – Virtual rendering of surface *S*)



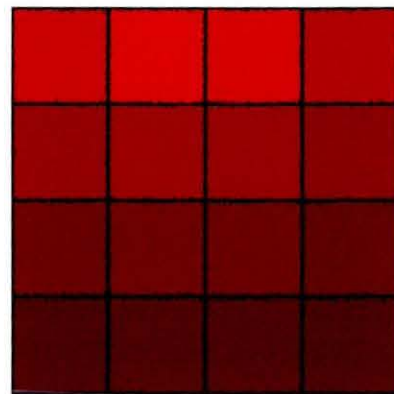
***Introducing a new surface***

A new surface,  $T$ , is introduced with original boundary coordinates (5,4,1), (5,2,1), (5,2,-1) and (5,4,-1). Surface  $T$ , thus, abuts surface  $S$  at  $90^\circ$ . The enclosure is used to re-illuminate the scene, as shown in figure 9.24, giving the result illustrated in figure 9.25.



(Figure 9.24 – Enclosure showing original surface  $S$  and virtual surface  $T$ )

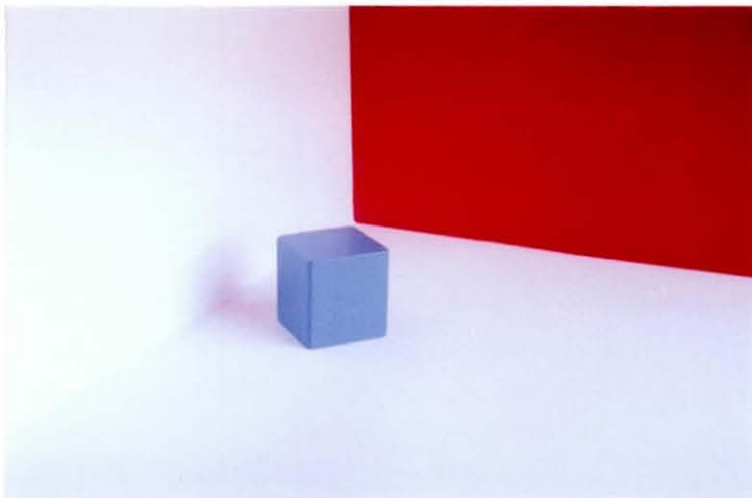
|        |        |        |        |
|--------|--------|--------|--------|
| 0.1953 | 0.2089 | 0.2042 | 0.1769 |
| 0.1447 | 0.1530 | 0.1487 | 0.1308 |
| 0.1019 | 0.1070 | 0.1040 | 0.0930 |
| 0.0655 | 0.0739 | 0.0720 | 0.0708 |



(Figure 9.25 – Resultant radiosities for surface  $T$ )

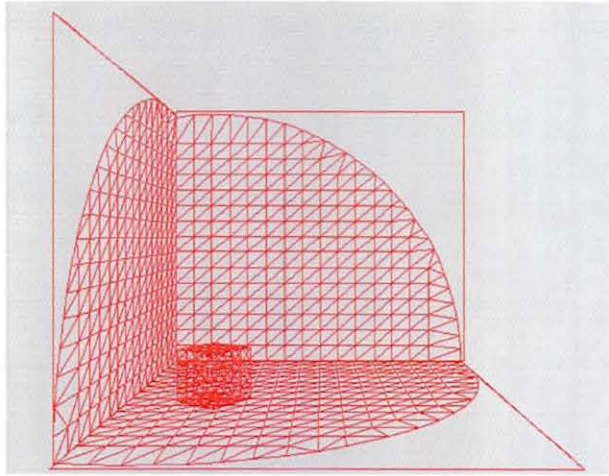
#### 9.6.4 Determining surface radiosities within a real scene

This section describes testing the illumination inference algorithm using a real scene. The test scene was constructed on a tabletop using cardboard, creating a floor and two walls, all perpendicular to one another. A child's building brick was placed within the scene, as shown in figure 9.26, and lit from a large window situated to the right-hand side of the table. The floor and one wall were white and the other wall; red. The brick was cyan. These colours were selected so that approximate reflectivities could be easily estimated. The white floor and left-hand wall are assumed to have RGB reflectivities of  $\text{RGB}(1.0,1.0,1.0)$ , the red back wall;  $\text{RGB}(1.0,0.0,0.0)$  and the cyan cube;  $\text{RGB}(0.0,1.0,1.0)$ .

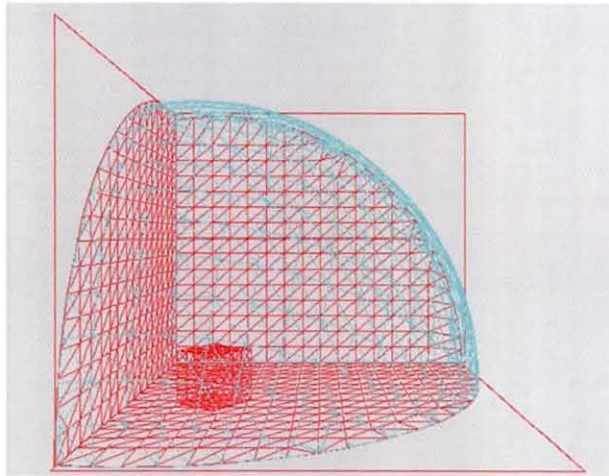


(Figure 9.26 – Test scene)

A tessellated computer model of the scene was created using coordinates measured using a millimetre rule. 3Dstudio [KINE03] was used to construct this model (figure 9.27) and to determine camera parameters. It was also used to create a tessellated sphere enclosing the environment. This sphere was sliced so that the volume below the floor or behind the walls was removed (figure 9.28). Triangle coordinates were exported for use in the illumination inference calculation and for the radiosity re-rendering step.



(Figure 9.27 – Tessellated reconstruction of test scene)

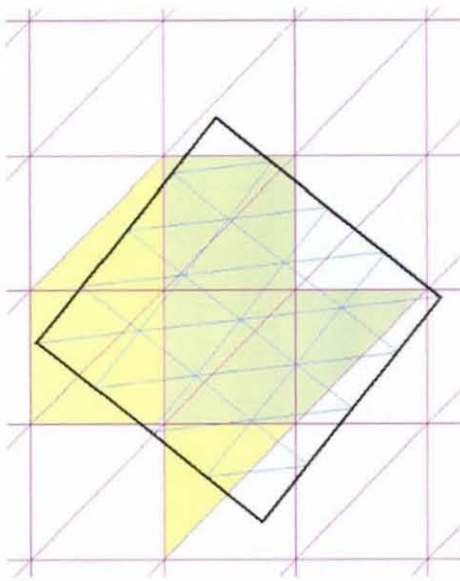


(Figure 9.28– Tessellated reconstruction of test scene  
with faceted enclosure shown in cyan)

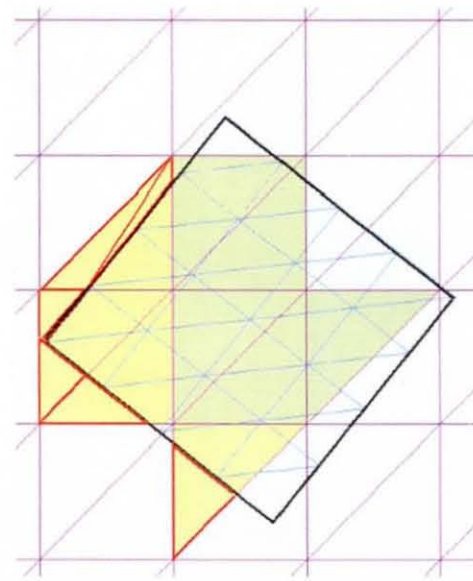
The total numbers of triangular facets were as follows.

|               |                    |
|---------------|--------------------|
| Back wall:    | 398 facets         |
| Left wall:    | 407 facets         |
| Floor:        | 466 facets         |
| Cube:         | 192 facets         |
| Enclosure:    | 563 facets         |
| <b>Total:</b> | <b>2006 facets</b> |

Where the cube is in contact with the floor, the facets can be ignored, as all remain completely occluded. Thus, the bottom face of the brick need not be considered. In the plan view (figure 9.29) the floor facets are shown as magenta and the brick's as cyan. The brick footprint is outlined in black for clarity. The green shaded floor facets are those that are completely obscured by the brick so can be ignored. The inter-facet form factors are calculated at facet centres, so the floor facets, shown coloured yellow, will not render correctly because their centres are obscured. To overcome this problem, these facets are further tessellated, as shown outlined red in figure 9.30. Partially obscured facets whose centre is not obscured will render correctly.



(Figure 9.29 – Plan view of brick and underlying floor to show obscured facets)

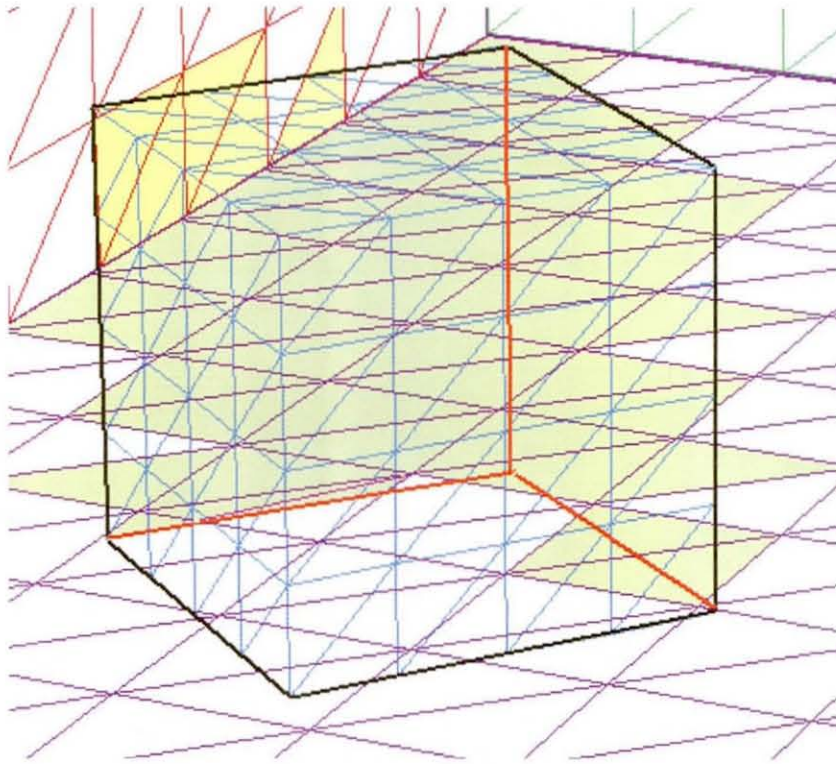


(Figure 9.30 – Plan view of brick and underlying floor showing facets, outlined in red, created where original facet centres are obscured)

The brick base and covered floor account for 41 facets. 6 new facets are created. Thus the total number of facets becomes:

$$2006 - 41 + 6 = 1971$$





(Figure 9.31 – Facets occluded by the brick.

Floor facets are shaded green and wall facets yellow)

There are 5 sidewall facets obscured from the viewer by the brick, shown shaded yellow in figure 9.31, and 47 further floor facets, shown shaded green. Only 3 sides of the brick are visible to the viewer, with 64 facets hidden, not including the base facets already discounted.

Therefore, in the whole scene there are:

- 1292 visible scene facets,
- 116 invisible scene facets, and
- 563 enclosure facets.



Thus, in equation 9.10:

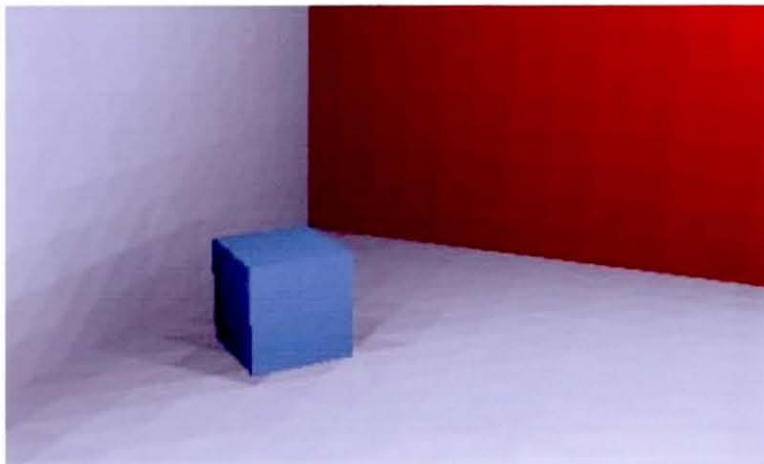
- $r = 1292$  (the number of visible surface patches - radiosity known),
- $s = 116$  (the number of invisible surface patches – radiosity unknown), and
- $t = 563$  (the enclosure facets - radiosity to be determined).

We therefore have  $(r + s) = 1408$  equations with  $(s + t) = 679$  unknowns.

Including artificial variables to account for an error term in each equation, matrix **A** becomes a  $1408 \times 2087$  matrix, and **x** a  $2087 \times 1$  vector.

To determine radiosity values  $B_r$  for the visible surfaces, original scene image pixel intensity values are taken to be directly proportional to their radiosity. This represents an approximation as the relationship between actual radiance (*power/area\*solid angle*) or radiosity (*power/area*) and pixel values is not known and actually depends in complex ways on characteristics of the imaging and digitising system.

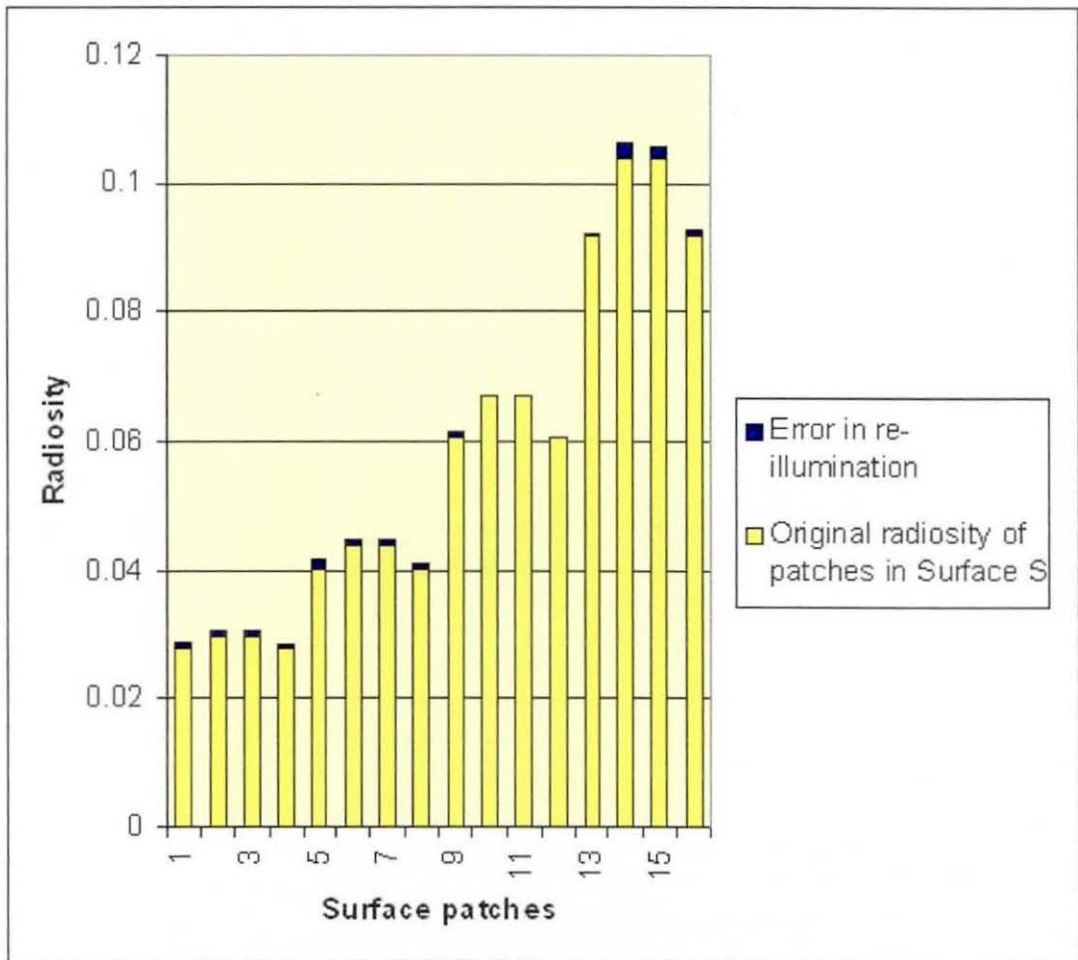
Form factors were calculated and the FSQP, Quadratic Programming Software, run to find an optimised solution for the enclosure facets. This was then used in re-rendering the scene as shown in figure 9.32.



(Figure 9.32 – Re-rendered scene)

### 9.7 Discussion

Applying the proposed illumination inference algorithm to re-illuminate a simple plane surface gives results mathematically close, and visually almost indistinguishable, from the original; see figure 9.33. In addition, augmenting the original scene with a new surface produces a visually plausible result.



(Figure 9.33. Bar plots showing original radiosities of the 16 patches of Surface  $S$  and the proportional error after re-illumination)

The real scene test also produces a visually plausible result. However, for this number of facets, it is not possible to verify that the FSQP implementation has found an optimised solution. The reliability of this implementation for solving such large-scale quadratic programming problems is suspect. Subsequent runs of the

implementation, with a second brick added, produced inconsistent results; therefore tests have not been conducted using more complex scenes.

Unfortunately, currently available quadratic programming implementations do not support the solution of very large-scale problems. The reasons are discussed by Gould and Toint [GOUL99] who, in a survey of SQP methods for large-scale non-linear programming, express the view that:

“In our opinion, the curious divergence between what logically should have happened in the 1980s, and what actually came to pass may be attributed almost entirely to a single factor: quadratic programming (QP) methods were not then capable of solving large problems. Witness the almost complete lack of software for solving large-scale quadratic programs even today, especially in view of the large number of available codes for the superficially similar linear programming problem.” (p.1)

However, they conclude:

“The majority of these are well-suited to large-scale problems.” (p.18)

which indicates that future implementations are likely to overcome the limitations encountered here.

## 10.0 Discussion and conclusion

### 10.1 Discussion

The last decade has seen considerable interest in systems that use computer-generated augmentation to enhance human perceptual experience of the real world. In particular, augmenting reality with computer-generated overlays, which has been the focus of this thesis, promises a wide range of new applications in diverse fields including archaeology, construction, design, education, entertainment, engineering and surgery. However, the visual channel is not the only human sense that can be computer enhanced. Augmenting reality through auditory and haptic senses is also the subject of current research.

For example, Meijer [MEIJ02] of Philips Research Laboratories is developing a system called *The vOICe*, which is designed to aid the blind by converting images from the real environment into sound. The user wears a small camera and its images are translated into sounds, a column of pixels at a time working from left to right. Two parameters are used to vary the sound generated. Pixels situated near the top of the picture are converted into high tones while those near the bottom are converted into low tones, and the brighter the pixel, the louder the sound, so a bright pixel near the top of the image would be high pitched and loud. After an entire scan, which takes about a second, the scan begins again. If the image changes, so does the pattern of sound. The system has received a lot of publicity and promises to be an effective low-vision aid.

With respect to the haptic channel, Hong et al. [HONG97] propose a system to aid endoscopic surgery. In their system, a potential field based on CT data is used to prevent the endoscope damaging the human colon as it is navigated. Another approach to haptic augmentation is provided by Mendoza et al. [MEND01] who propose a system which allows deformable virtual objects to be touched using a physical simulation. Even more recently, Nojima et al. [NOJI02] have developed *SmartTool*, which uses sensors to measure some property of the real environment and

then translates this into haptic sensation for the user. An envisaged future application for *SmartTool* is in surgery such that when a surgical instrument is near a vital tissue that should not be damaged, the system would provide force feedback to avoid harm occurring.

Whichever human sense is to be augmented, the fundamental issue remains the same. Sufficient information must be obtained from the real environment so that the augmentation can be appropriately synthesised and integrated with it. As far as augmenting the visual channel is concerned, the processing demands in this regard depend on the application. For example, a system that overlays textual information on views of the real world may only need a sparse model of environment geometry and no knowledge of its lighting. However, in a design visualisation application, a very complete real-world model may be needed. To summarise, in an ideal system offering completely seamless visual integration of the real and the virtual, we need to be able to:

- acquire and maintain a model of the real-world geometry so we can use the relative depth and shape of virtual and real entities to determine appropriate occlusion, shading and shadow interactions,
- obtain a model of the material properties of real-world objects so we can render a composite scene using a common illumination model,
- track the user's position, orientation and line-of-sight with sufficient spatial and temporal accuracy over required distance,
- generate synthetic augmentations that are realistically rendered as well as spatially and temporarily registered with the real world,
- display the augmented scene to the user in a way that does not spatially or temporarily distort reality, and does not interfere with the users ability to carry out tasks.



### ***10.1.1 Modelling real-world geometry***

Acquiring and maintaining a model of real-world geometry has been discussed in chapter 8. For some applications a computer model of the real environment may already exist, although a built environment will almost inevitably differ in some respects from its design model and, unfortunately, no current techniques provide an ideal solution for achieving real-time, accurate 3D scene reconstruction. Laser range scanning produces the best results. However, improvements in passive camera-based techniques are the subject of ongoing research and future advances seem likely.

### ***10.1.2 Modelling material properties***

Obtaining information about the material properties of real-world objects remains problematic. In generating a composite rendering of a scene, we need to be able to re-compute the interaction of light with surfaces in the scene and to do this we need to know the reflectance properties of the surfaces. A complete specification of a surface requires determination of its bidirectional reflectance distribution function (BRDF), which describes the proportion of light reflected from any given incident direction to any given view direction. Currently, this can only be accurately measured in laboratory conditions. Most previous work on surface reflectance measurement has used carefully controlled laboratory lighting [SATO97][TOMI00], although Yu et al. [YU99] have been able to determine reflectances in a room given a set of photographs of all surfaces and knowledge of the light sources. Research is ongoing [eg DROR03] to provide ways of determining the reflective properties of an object's surface from a single image.

If the complete environment geometry and material properties are known, including light sources, then we have the information needed for a common illumination model. However, in the more probable situation, where information is from a restricted field of view, ways are needed of acquiring an appropriate illumination model with which to light the composite scene. As discussed in chapter 9, current solutions rely on prior measurement and a new approach is proposed as part of this thesis.

### ***10.1.3 Maintaining registration***

Accurate tracking, sometimes over large distance, is required for augmented reality. The range of available approaches has been discussed in chapter 3. Unfortunately, methods with long range, such as GPS, lack accuracy, and more accurate techniques, such as optical tracking, lack range. (GPS typically has a resolution of between 1 to 10 metres, depending on whether Differential GPS is used, while optical trackers can have sub-millimetre accuracy, but a range of only up to about 6 metres. See Rolland et al. [ROLL00] for a detailed comparison of tracking method specifications.) Future practical strategies may involve hybrids that combine the advantages of a mixture of existing approaches.

Current techniques are able to generate synthetic augmentations that are realistically rendered, but scene generation can be slow for photo-realistic rendering. Improvements in algorithm efficiency and faster computational platforms will further reduce temporal delays.

### ***10.1.4 Displaying real-virtual composites***

Improvements in displays for optical-see-through augmented reality have made little progress. The popular Sony Glasstron PLM-S700E [SONY03], which had a see-through mode, is no longer available. However, a 1024 x 768 pixel resolution monochrome display, ProView XL40/50 STm, is available from Kaiser Electro-Optics Inc. [KAIS03], and Microvision Inc. [MICR03] now manufacture a monocular colour display giving 800 x 600 pixel resolution. This latter system uses virtual retinal scanning technology [PRYO98], which scans a VGA image directly onto the retina using a low-power laser, and was developed in the Human Interface Technology Laboratory (HITL) at the University of Washington. None of these displays allow occlusion of real-world objects by virtual. This, and the related issue of displaying real-virtual illumination effects are addressed by the display arrangements proposed and demonstrated in chapter 7. Future displays seem likely to make use of microelectromechanical machines (MEMs) such as the digital

micromirror device (DMD) also discussed in chapter 7. Wearable devices, using this technology, will present considerable challenge and have yet to be developed.

## **10.2 Conclusion**

The concept of virtual reality is well known, although surprisingly few applications have progressed beyond the research laboratory and into popular use, perhaps the most notable exceptions being for training simulations and engineering prototyping. There are a number of possible reasons for this, including the sometimes limited return on investment in specialised equipment and the difficulty in producing useful virtual worlds.

Immersive virtual reality insulates its user from the real environment, but it is through interaction with real-world objects that most useful tasks are accomplished. Augmented reality systems allow users some of the benefits of a virtual environment, while maintaining perceptual input from the real surroundings. In particular, virtual reality allows users to perceive computer-based information without the spatial constraint of a limited-size display screen. By superimposing artificial images on the real environment, augmented reality systems can embellish it with additional information or enable non-existent and invisible objects to be seen. This promises to fulfil new applications in a wide range of fields including surgery, design, archaeology, engineering, construction, education and entertainment. However, for the promise of augmented reality to reach fruition, there are some challenging issues to be resolved.

An important issue relates to achieving accurate registration of virtual and real worlds. Not only should the computer-generated elements align correctly in real space, but also in time. Correct spatial registration depends on system calibration, as well as accurate tracking of the user and of any real objects that are to be manipulated. Appropriate temporal registration can only be achieved if delays in the graphic update are sufficiently small for the graphic display to keep pace with user movements. The accuracy of registration required depends on the application. In

some domains, such as annotating real-world objects with a textual description, exact registration may not be crucial. However, in others such as surgery, sub-millimetre accuracy may be essential. Not surprisingly, registration issues have attracted considerable research attention and future solutions for augmented reality are likely to involve hybrid tracking in which the respective advantages of different tracking technologies can be exploited.

Accurate spatial registration is not only required with respect to lateral positioning but also in depth. A limiting problem with existing optical-see-through displays is that they are incapable of displaying a full range of depth cues. Most significantly, their optics always produce ‘ghost-like’ virtual overlays that are unable to occlude real background and hence cannot produce interposition depth cueing. Neither are they able to modify the real-world view in the ways required to produce convincing common illumination effects such as virtual shadows, which also are potentially useful cues to depth.

How useful these cues are to the estimation of depth in augmented realities may seem intuitive but has never before been quantified. The experiment described in chapter 5 of this thesis presents a method and apparatus for estimating virtual depth. It establishes that appropriate interpositioning is essential to accurate estimation of depth in augmented realities, and that the presence of shadows provides an important refining cue.

Creating appropriate real-virtual shadows exemplifies another set of technological problems that frustrate the successful exploitation of augmented reality. These issues relate to finding ways to appropriately simulate the interactions that should occur between real and virtual objects. The interactions could be physical contact or the complex interactions of light between objects in proximity with one another. Solutions to both of these issues are dependent on having a computer-based model of the real environment as well as the virtual. To deal with simple collisions, an approximate bounding-box model of real-world geometry may suffice. However,

modelling more subtle interactions, such as the depression of a real cushion under the 'weight' of a virtual object, requires knowledge of material properties as well as a way to visually alter the appearance of the real-world scene. Similarly, light interactions between real and virtual objects present considerable difficulty. For example, virtual objects added to a real environment should be lit in the same way as real objects and should cast shadows on reality. Also, virtual light sources should be capable of dissipating real shadows, and there may be objects that are not within view but that affect the illumination of a scene.

Full resolution of these problems requires a complete model of real-world geometry, a model of the material properties of all real-world objects, and a model of the real-world illumination. Not only must suitable graphic augmentations be generated but also many real-virtual interactions involve modification of the real-scene itself. Such alteration is possible using video-see-through and monitor-based displays, as both the virtual and the real are electronically composited. This means that the real-scene view can be modified in any desired fashion before presentation to the user. However, this is not the case with optical-see-through displays where a direct view of reality is maintained and cannot currently be modified on a localised basis.

Nevertheless, there are a number of potential advantages in using optical-see-through displays, depending on application. For instance, they do not necessarily present a restricted field of view, or reduce the resolution with which the real scene is viewed, neither of which is true of video-see-through displays. Also, they do not produce the displacement due to the offset between the user's eyes and the cameras that is characteristic of video-see-through systems. First-hand experience of the real world is maintained, and optical-see-through systems can be made fail-safe so that vision is not completely disrupted in the event of power failure. What was needed, therefore, was a way of retaining the advantages of optical-see-through augmented reality while providing facility to modify the user's view of real-world entities.



This thesis presents a solution to this requirement by extending the concept of a transparency alpha-channel to direct views of the real world. Three approaches to achieving this are presented. The first uses a transmissive mask to actively filter the real scene. The second employs digital projection to illuminate the scene in a selective way. The third technique utilises digital micromirror devices to selectively reflect elements from the real or virtual scenes. The generalised theory of the approach is described mathematically in chapter 6 and algorithms developed to automate generation of alpha-mask and overlay images, including a holdout matte for areas where occlusion is required.

The transmissive approach has been effectively demonstrated using a LCD panel for static scenes and for a simple animated world. The limitations of employing LCD panels as masks were found to be high attenuation of light and also the problem of trying to simultaneously accommodate the user's eyes to the mask and the real scene. The latter issue has subsequently been addressed using lenses to bring the real scene into focus in the same plane as the panel.

For situations where real-world lighting can be controlled, the projective approach to alpha-channel masking may be suitable and this has been demonstrated applied to the virtual restoration of a polychrome carving. It was proposed that an ideal solution to real-virtual compositing may lie in the use of digital micromirror devices, and this thesis has described how alpha-channel masking may be implemented using these reflection devices. Considerable work is needed in developing this concept if the resulting display hardware is to be wearable.

It is possible that better transmissive display technologies will emerge in the future. Amongst the newest technologies are organic light-emitting diodes (OLED). These are formed from evaporated thin films of stable organic materials that emit light of various colours when a voltage is applied. Compared to LCDs, they offer a number of advantages such as lighter and thinner devices as well as lower power

consumption [KIMM02]. However, it remains to be seen whether such devices will prove suitable for transmissive alpha-channel masking.

Whatever, the display technology ultimately employed, the concept of alpha-channel masking real-world views, proposed in this thesis, remains a fundamental requirement for seamless visual integration in optical-see-through augmented reality systems. Also of fundamental importance to full and convincing integration, is the acquisition of sufficient knowledge of the real scene to enable appropriate real-virtual light interaction effects to be realised. Ultimately this would necessitate real-time determination of scene geometry, material reflectance properties, and a common illumination model to enable correct relighting after augmentation. All of these pose significant challenges.

As for as obtaining a common illumination model is concerned, all current approaches require either prior knowledge of the light sources illuminating the real scene, or involve inserting some kind of probe into the scene with which to determine real light source position, shape, and intensity. This thesis presents an alternative approach that infers a plausible illumination model from a limited view of the scene. This offers the potential advantage of using a single camera view to extract a common illumination model without prior physical probing of the scene. The method envelops the scene within a virtual faceted enclosure for which an optimised radiosity solution is then found. This optimisation problem is non-linear and large and, although it has been demonstrated for simple cases, cannot be extended to more complex scenes without the future development of software implementations for solving large quadratic programming problems.

The main contributions made by this thesis to the scientific community comprise:

- experimental verification of the importance of occlusions and shadows in estimating depth in augmented reality scenes (described in chapter 5),
- the mathematical basis for creating and applying a real-world alpha channel to achieve occlusion and shadow effects in optical-see-through displays (described in chapter 6),
- improved optical-see-through display designs, allowing a real-world alpha channel to be realised (described in chapter 7), and
- a new strategy for inferring a common illumination model from a scene, without knowledge of original light source parameters (described in chapter 9).

The author has published work on spatial and temporal registration in augmented reality [TATHA97][KALA98]. Work has also been published relating to the alpha-channel masking concept [TATH99a][TATH99b] for which a provisional patent application was filed, No. GB9807107.9. (See appendix D.) Work on the common illumination inference algorithm has also been presented and published [TATH99c]. All associated papers are included as appendix E.

The future of augmented reality appears promising and research interest has developed rapidly since wider attention was drawn to its possibilities by the July 1993 issue of the *Communications of the ACM*, which made augmented reality its theme. Having produced this thesis as a part-time researcher, with the work being interrupted due to personal circumstances, the author has been a close monitor of progress made since that time. Despite the frenetic feelings of pace that competitive concerns can raise, overall progress has been modest and the key problems remain. However, it is hoped that the contribution made by this thesis moves us a bit closer to the fulfilment of more tractable augmented reality applications.



**Bibliography**

[AEM03] AEM Design, 3754 LaVista Rd., Suite 250, Tucker, GA 30084, USA (www.aemdesign.com).

[AHLE95] Ahlers K, Kramer A, Breen D, Chevalier PY, Crampton C, Rose E, Tuceryan M, Whitaker R and Greer D. *Distributed augmented reality for collaborative design applications*. Proc. Eurographics '95 Conf., Maastricht, Netherlands, (August 1995), pp3-14.

[ALIA97] Aliaga DG. *Virtual Objects in the Real World*. Communications of the ACM, vol 40, no. 3, (1997), pp49-54.

[AMES25] Ames A. Jr. *The illusion of depth from single pictures*. Journal of the Optical Society of America, v10, (1925), pp137-148.

[ANDE94] Anderson BL and Nakayama K. Towards a general theory of stereopsis: Binocular matching, occluding contours and fusion. *Psychological Review*, 101, (1994), pp414-445.

[ARVI03] ARVIKA Project, (www.arvika.de), (2003).

[AZUM93] Azuma R. *Tracking Requirements for Augmented Reality*. Communications of the ACM, 36, 7 (July 1993), pp50-51.

[AZUM95] Azuma R. *Dissertation: Predictive Tracking for Augmented Reality*. UNC Chapel Hill Dept. of Computer Science technical report TR95-007 (February 1995).

[BAJU92] Bajura M, Fuchs H and Ohbuchi R. *Merging virtual objects with the real world: seeing ultrasound imagery within the patient*, *Computer Graphics*, 26 (2), (1992), pp203-210.

[BARN86] Barnes J and Hut P. *A hierarchical  $O(N \log N)$  force-calculation algorithm*. *Nature*, 324(4), (December 1986), pp446-449.

[BAUD93] Baudel T and Beaudouin-Lafon M. *Charade: Remote Control of Objects using FreeHand Gestures*. Communications of the ACM, vol 36, no. 7, (1993), pp. 28-35.

[BESL92] Besl P and McKay N. *A Method for Registration of 3D Shapes*. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2), (1992), pp239-255.

[BIOC92] Biocca F. *Will simulator sickness slow down the diffusion of virtual environment technology?* *Presence*, 1 (3), (1992), pp258-264.



- [BREE95] Breen D, Rose E, and Whitaker R. *Interactive occlusion and collision of real and virtual objects in augmented reality*, Technical Report ECRC-95-02, ECRC, Munich, Germany, (1995).
- [BRUC96] Bruce V, Green PR, and Georgeson MA. *Visual Preception. Physiology, Psychology, and Ecology*, 3rd Edition, Psychology Press, (1996).
- [CAME91] Cameron J. (director) *Terminator 2: Judgement Day*, Columbia Tristar, James Cameron, 1991.
- [CATM74] Catmull E. *A subdivision algorithm for computer display of curved surfaces*. PhD thesis, University of Utah, (1974).
- [CAUD92] Caudell T and Mizell D. *Augmented reality: An application of heads-up display technology to manual manufacturing processes*. Proc. Hawaii Int. Conf. System Sciences, (Jan 1992), pp659-669.
- [COHE85] Cohen MF and Greenberg DP. *The Hemi-Cube - A Radiosity Solution for Complex Environments*. Computer Graphics 19, No. 3, SIGGRAPH '85 Proceedings, (July 1985), pp31-40.
- [COHE86] Cohen MF, Greenberg D., Immel DS, and Brock DJ. *An Efficient Radiosity Approach for Realistic Image Synthesis*, IEEE Computer Graphics and Applications, (March 1986), pp26-35.
- [COHE88] Cohen MF, Chen SE, Wallace JR, and Greenberg DP. *A Progressive Refinement Approach to Fast Radiosity Image Generation*, Computer Graphics 22 No.4, SIGGRAPH '88 Proceedings, (August 1988), pp75-84.
- [COOK84] Cook RL, Porter T and Carpenter L. *Distributed ray tracing* Computer Graphics, 18(3), (1984), pp137-145.
- [CORR02] Corrêa WT, Fleishman S and Silva CT. *Towards Point-based Acquisition and Rendering of Large Real-world Environments*. Proceedings of the 15th Brazilian Symposium on Computer Graphics and Image Processing, (2002), pp59-66.
- [CURL96] Curless B and Levoy M. *A Volumetric Method for Building Complex Models from Range Images*. In ACM Computer Graphics Proceedings, SIGGRAPH, New Orleans, (1996), pp303-312.
- [CULB00] Culbertson WB, Malzbender T, and Slabaugh G. *Generalized voxel coloring*. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice* (Proc. Int. Workshop on Vision Algorithms) , volume 1883 of Lecture Notes in Computer Science, Springer-Verlag, (2000), pp100-115.

- [CUTT97] Cutting J. *How the eye measures reality and virtual reality*. In Behavior Research Methods, Instruments & Computers, 29(1), (1997), pp27-36.
- [DEBE98] Debevec PE. *Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography*. In M. Cohen, editor, SIGGRAPH '98 Conference Proceedings, Annual Conference Series, (1998), pp189-198.
- [DERE72] Deregowski J. *Pictorial Perception and Culture*, Scientific American, Vol. 227, No.5, (Nov 1972), pp82-88.
- [DRAS93] Drascic D, Grodski JJ, Milgram P, Ruffo K, Wong P, and Zhai S. *ARGOS: A Display System for Augmenting Reality*, ACM SIGGRAPH Technical Video Review, Volume 88: InterCHI '93 Conf on Human Factors in Computing Systems, (1993), p521.
- [DRET98] Drettakis G, Robert L, and Bougnoux S. *Interactive common illumination for computer augmented reality*. In J. Dorsey and P. Slusallek, editors, *Rendering Techniques '98*. 8th EG workshop on Rendering, St. Etienne, France, (June 1997), pp45-56.
- [DROR03] Dror R and Adelson T. *Estimating Surface Reflectance from Images*, Artificial Intelligence Laboratory, Massachusetts Institute Of Technology, Cambridge, Massachusetts 02139.
- [EDWA03] Edwards PJ, King AP, Maurer Jr CR, de Cunha DA, Hawkes DJ, Hill DLG, Gaston RP, Fenlon MR, Juszczek A, Strong AJ, Chandler CL, and Gleeson MJ. *Design and Evaluation of a System for Microscope-Assisted Guided Interventions (MAGI)*, IEEE Transactions on Medical Imaging, in press, (2003).
- [ELLI94] Ellis SR and Bucher UJ. *Distance Perception of Stereoscopically Presented Virtual Objects Optically Superimposed on Physical Objects by a Head Mounted See-Through Display*. Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society, Nashville, Tennessee, (1994), pp1300-1304.
- [ELLI96] Elliot D. *Understanding GPS: principles and applications*. Ed. E Kaplan, Artec House Publishing, (1996).
- [EMME81] Emmert E. *Grossenverhältnisse der Nachbilder*. Klinische Monatsblätter für Augenheilkunde 18, (1981), pp443-450.
- [FEIN93] Feiner, S, MacIntyre B, and Seligmann D. *Knowledge-based augmented reality*. Communications of the ACM, 36(7), (July 1993), pp52-62.

- [FEIN95] Feiner S, Webster A, Krueger T, MacIntyre B, and Keller E. *Architectural Anatomy*. Presence: Teleoperators and Virtual Environments, 4(3), (1995), pp318-325.
- [FERR72] Ferris SH. Motion parallax and absolute distance. *Journal of Experimental Psychology*, 95, (1972), pp258-263.
- [FLEI00] Fleishman S, Cohen-Or D, and Lischinski D. *Automatic Camera Placement for Image-based Modeling* *Computer Graphics Forum*, 19(2), (2000), pp100-110.
- [FOUR94] Fournier A. *Illumination Problems in Computer Augmented Reality*, Journée INRIA, Analyse/Syntaxe D'Images, (Jan 1994), pp1-21.
- [FUCH96] Fuchs H, State A, Pisano ED, Garrett WF, Hirota G, Livingston MA, Whitton MC and Pizer SM. *Towards Performing Ultrasound-Guided Needle Biopsies from Within a Head-Mounted Display*. Proc. Visualization in Biomedical Computing, Hamburg, Germany, (Sept 1996), pp591-600.
- [FUCH98] Fuchs, H, Livingston MA, Raskar R, Colucci D, Keller K, State A, Crawford JR, Rademacher P, Drake SH, and Meyer AA. *Augmented Reality Visualization for Laparoscopic Surgery*. Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI) '98, Cambridge, MA, (Oct, 1998), pp934-943.
- [FURN69] Furness T. *Helmet-Mounted Displays and Their Aerospace Applications*; National Aerospace Electronics Conference, Dayton, OH, (May 1969).
- [GENN92] Gennery DB. *Visual tracking of known three dimensional objects*, *International Journal of Computer Vision*, 7(3), (1992), pp243-270.
- [GEOR03] Georgia State University.  
<http://hyperphysics.phy-astr.gsu.edu/hbase/vision/cie.html>, 2003.
- [GIBS50] Gibson JJ. *The Perception of the Visual World*, Houghton Mifflin, Boston, (1950).
- [GLAS89] Glassner A. *An Introduction to Ray Tracing*, Academic Press, (1989).
- [GORA84] Goral CM, Torrance KE, Greenberg DP and Battaile B. *Modeling the Interaction of Light Between Diffuse Surfaces*, Proceedings of SIGGRAPH '84, Computer Graphics, Annual Conference Series, (1984), pp213-222.
- [GOUL99] Gould NIM. and Toint PL. *SQP methods for large-scale nonlinear programming*. Technical Report RAL-TR-1999-055, Computational Science and Engineering Department, Atlas Center, Rutherford Appleton Laboratory, (1999).

- [GRAH65] Graham CH. *Vision and Visual Perception*, Wiley, (1965).
- [GREE99] Greenspan MA and Boulanger P. *Efficient and Reliable Template Set Matching for 3D Object Recognition*. Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling, (1999), pp230-239.
- [GREG63] Gregory RL. *Distortion of visual space as inappropriate constancy scaling*. Nature, 199, (1963), pp678-91.
- [GREG77] Gregory RL. *Eye and Brain - The Psychology of Seeing*, Weidenfeld and Nicolson, (1997).
- [GUNK95] Gunkel AR., Freysinger W., Thumfart WF. & Truppe, MJ. *Application of the ARTMA image-guided navigation system to endonasal sinus surgery*, Proc. (Computer Assisted Radiology) CAR95, June 21-24, Berlin,(1995), pp1146-1151.
- [GYRA03] Gyration, Inc., 12930 Saratoga Avenue, Bldg.C, Saratoga, CA 95070.
- [HABE80] Haber R. and Hershenson M. *The Psychology of Visual Perception* (2nd edition). Holt, Rinehart and Winston, (1980).
- [HAKI98] El-Hakim S, Brenner C and Roth G. *An Approach to create Virtual Environments using Range and Texture*. In ISPRS International Symposium on Real Time Imaging, (June 1998), pp331-338.
- [HANR91] Hanrahan P, Salzman D, and Auperle L. *A rapid hierarchical radiosity algorithm*. Computer Graphics, 25(4), (August 1991), pp197-206.
- [HART97] Hartley P and Sturm R. Triangulation. *Computer Vision and Image Understanding*, (1997).
- [HIIP 97] Hiipakka J, Hänninen R, Ilmonen T, Napari H, Lokki T, Savioja L, Huopaniemi J, Karjalainen M, Tolonen T, Välimäki S and Takala T. *Virtual Orchestra Performance*. SIGGRAPH Visual Proceedings, (Aug 1997), p81.
- [HITL003] Human Interface Technology Laboratory (HIT Lab), University of Washington, Seattle, WA 98195-2142, USA.
- [HOLL99] Hollerer T, Feiner S, Terauchi T, Rashid G, and Hallaway D. *Exploring mars: Developing indoor and outdoor user interfaces to a mobile augmented reality system*. Computers and Graphics, 23(6), (Dec 1999), pp779--785.
- [HONG97] Hong L, Muraki S, Kaufman A, Bartz D, and He T. *Virtual Voyage: Interactive Navigation in the Human Colon*, Proceedings of SIGGRAPH '97, (1997), pp27-34.

- [HORN70] Horn BKP, *Shape from Shading, a Method for obtaining the shape of a smooth opaque object from one view* PhD Thesis, Dept of Electrical Engineering MIT, (1970).
- [HORN89] Hornbeck LJ. *Deformable-Mirror Spatial Light Modulators*. Spatial Light Modulators and Applications III, SPIE Critical Reviews, Vol. 1150, (August 1989), pp86-102.
- [JAU91] Jau B. *Technical support Package on Anthropomorphic Remote Manipulator for NASA*. TECH BRIEF, 15(4) from JPL Invention Report (Report No.NPO-17975/7222). Pasadena, CA: JPL Technology Utilization Office, (1991).
- [JOHA73] Johansson G. Monocular movement parallax and near-space perception. *Perception*, 2, (1973), pp136-145.
- [JOHN93] Johnston EB, Cumming BG, and Parker AJ. *Integration of depth modules: Stereopsis and texture*. *Vision Research*. 33, 5/6, (1993), pp813-826.
- [KAIS03] Kaiser Electro-Optics, Inc., 2752 Loker Avenue West, Carlsbad, CA 92008.
- [KALA98] Kalawsky RS and Tatham EW. *Effects of Spatial and Temporal Mis-registration in Augmented Virtual Environments*, Proceedings of the 16th Annual Conference, Eurographics UK, Leeds, (March 1998), pp127-134.
- [KIMM02] Kimmel J, Hautanen J and Levola T. *Display Technologies for Portable Communication Devices*. Invited Paper in Proceedings of the IEEE, Vol. 90, No.4, (April 2002), pp581-590.
- [KINE03] Kinetic 3DStudio MAX, Autodesk, Inc., 111 McInnis Parkway, San Rafael, CA 94903, USA.
- [KIYO00] Kiyokawa K, Kurata Y and Ohno H. An Optical See Through Display for Mutual Occlusion of Real and Virtual Environments. Proc. International Symposium Augmented Reality, ISAR 2000, Los Alamitos, California, (2000), pp60-67.
- [KNOW77] Knowlton KC. *Computer displays optically superimposed on input devices*, Bell System Technical Journal, vol. 56, (1977), pp367-83.
- [KOEN94] Koenderink JJ, van Doorn AJ, and Kappers AM. *On so-called paradoxical monocular stereoscopy*. *Perception*, 23(5), (1994), pp583-94.
- [KRUE91] Krueger M. *Artificial Reality II*. Addison-Wesley, New York, (1991).



- [KRUE85] Krueger M, Gionfriddo T, and Hinrichsen K. *Videoplace - an artificial reality*. In Proceedings of ACM CHI'85 Conference on Human Factors in Computing Systems. ACM Press, (1985), pp35-40.
- [KUTU99] Kutulakos S and Seitz K. *A Theory of Shape by Space Carving*. In International Conference on Computer Vision, (1999), pp307-314.
- [LAMB60] Lambert JH. *Photometria sive de mensura de gradibus luminis, colorum umbrae*. Eberhard Klett, (1760).
- [LAND95] Landy MS, Maloney LT, Johnston EB & Young M. Measurement and modeling of depth cue combination, *Vision Research*, 35, (1995), pp389-412.
- [LAUR99] Laurentini A. *The Visual Hull of Curved Objects*. Proc. IEEE Int. Conf. on Computer Vision, (1999), pp. 356-361.
- [LAWR97] Lawrence C, Zhou JL, and Tits AL. *CFSQP Version 2.5: A C Code for Solving Constrained Nonlinear Optimization Problems, Generating Iterates Satisfying All Inequality Constraints*, Electrical Engineering Dept and Institute for Systems Research, University of Maryland, College Park, MD 20742, (1997).
- [LOGI91] Logitech (1991) press kit.
- [MAYB93] Maybank SJ. *Theory of Reconstruction from Image Motion*. Springer-Verlag, Berlin, (1993).
- [MCMA97] McManners H. *The Sunday Times*, 2nd February, (1997), p5.
- [MEIJ02] Meijer P, *Seeing with Sound for the Blind: Is it Vision*, Tuscon 2002 Conference: "Toward a Science of Consciousness", Tuscon Arizona, (April 2002).
- [MEND01] Mendoza CA, Laugier C, *Realistic Haptic Rendering for Highly Deformable Virtual Objects*, Proceedings of 2001 IEEE Virtual Reality Conference, pp257-263, 2001.
- [METZ93] Metzger PJ. *Adding reality to the virtual*, Proc. IEEE Virtual Reality Annual Int. Symp., September 18-22, Seattle, WA, (1993), pp7-13.
- [MICR03] Microvision Inc., 19910 North, Creek Parkway, Bothell, WA , 98011-3008 USA.
- [MILG93] Milgram P, Zhai S, Drascic D, Grodski JJ. *Applications of Augmented Reality for Human-Robot Communication*. Proc. IROS'93: Int'l Conf. on Intelligent Robots and Systems, Yokohama Japan, (July 1993), pp14.

- [MILG94] Milgram P, and Kishino F. *A Taxonomy of Mixed Reality Visual Display*, IEICE Transactions on Information Systems special issue on Networked Reality, vol.E77-D, no.12, (Dec 1994), pp1321-1329.
- [MILG95] Milgram P, Drascic D, Grodski J, Restogi A, Zhai S, and Zhou C. *Merging real and virtual worlds*. Proc of IMAGINA '95, Monte Carlo, (Feb 1995), pp221-230.
- [NAKA86] Nakamae E, Harada K, Ishizaki T, and Nishita T. *A montage method: The overlaying of the computer generated images onto a background photograph*. In D. C. Evans and R. J. Athay, editors, Computer Graphics (SIGGRAPH '86 Proceedings), volume 20, (Aug 1986), pp207-214.
- [NITZ77] Nitzan D, Brain A, and Duda R. *The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis*. Proceedings of the IEEE, 65(2), (2000), pp206-220.
- [NOE94] Noe P, and Zabaneh K. *Relative GPS*. IEEE Position Location and Navigation Symposium, (1994), pp586-590.
- [NOJI02] Nojima T, Sekiguchi D, Inami M and Tachi S. *The SmartTool: A system for augmented reality of haptics*, Proceedings of IEEE Virtual Reality, (Mar 2002), pp67-72.
- [OCKE98] Ockerman, JJ, Thompson JC, and Najjar, LJ. *Wearable computers for performance support: Initial feasibility study*. Personal Technology, 1(4), (1998), pp251-259.
- [OHSH98] Ohshima, T, Satoh K, Yamamoto H and Tamura H. *AR2 Hockey: A Case Study of Collaborative Augmented Reality*. Proc. IEEE VRAIS '98, (1998), pp.268-275.
- [PARR89] Parr J. *Introduction to Ophthalmology (3rd Edition)* Oxford University Press, (1989).
- [PHON75] Phong B-T. *Illumination for Computer Generated Pictures* Communications of the ACM 18(6), (June 1975), pp311-317.
- [PICK89] Pickwell D. *Binocular Vision Anomalies (2nd edition)* Butterworths, (1989).
- [POLH03] Polhemus Inc. 40 Hercules Drive, PO Box 560 Colchester, VT 05446.
- [PORT84] Porter T and Duff T. *Compositing Digital Images*, Computer Graphics, Vol 18, No 3, SIGGRAPH'84 Conference Proceedings, (Jul 1984), pp253-259.

- [POTT67] Potter JH. *Handbook of the Engineering Sciences*. D. Van Nostrand Company, Inc., (1967).
- [PRIN03] Princeton Video Image Inc., Corporate Headquarters, 15 Princess Road, Lawrenceville, NJ 08648, (2003).
- [PRYO98]. Pryor HL, Furness TA and Viirre E. *The Virtual Retinal Display: A New Display Technology Using Scanned Laser Light*. In Proceedings of Human Factors and Ergonomics Society, 42nd Annual Meeting, 1998), pp157.
- [RASK99] Raskar R, Welch G, and Chen W. *Tabletop spatially augmented reality : Bringing physical models to life using projected imagery*. Second Int Workshop on Augmented Reality (IWAR'99), San Francisco, CA, (Oct 1999), pp 64-71.
- [REKI96] Rekimoto J. *TransVision: A hand-held augmented reality system for collaborative design in Virtual Systems and Multi-Media*. Proc. VSMM'96, Gifu, Japan, (Sept 1996), pp18-20.
- [REIT01] Reitmayr R, and Schmalstieg D. *Mobile Collaborative Augmented Reality*. Proc. ACM and IEEE International Symposium on Augmented Reality (ISAR'01), New York, (2001), pp114123.
- [RITT77] Ritter M. *Effect of Disparity and Viewing Distance on Perceived Depth*, Perception and Psychophysics, 22(4), (1977), pp400-407.
- [ROCK84] Rock I. *Perception*, Scientific American Library, (1984).
- [ROGE79] Rogers BJ. and Graham ME. *Motion parallax as an independent cue for depth perception*. Perception 8 (1979), pp125-134.
- [ROLL00] Rolland JP, Davis LD, and Baillet Y. *A survey of tracking technology for virtual environments*, in Fundamentals of Wearable Computers and Augmented Reality. (Chapter 3) Ed. Barfield and Caudell (Mahwah, NJ), (2000).
- [ROSC84] Roscoe SN. *Judgements of size and distance with imaging displays*. Human Factors, 26, (1984), pp617-629.
- [ROSC93] Roscoe SN. *The eyes prefer real images*. In Ellis SR, Kaiser MK, and Grunwald AJ (Eds.), Pictorial communication in virtual and real environments. (2nd ed.). London: Taylor and Francis, (1993), pp577-585.
- [ROSE95] Rose E, Breen D, Ahlers K, Crampton C, Tuceryan M, Whitaker R, and Greer D. *Annotating real-world objects using augmented reality*. Proc. of Computer Graphics: Developments in Virtual Environments International 95, (June 1995), pp357-370.

- [ROTH99] Roth G. *Registering Two Overlapping Range Images*. Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling, (1999), pp191-200.
- [RUTI94] Rutishauser M., Stricker M and Trobina M. *Merging Range Images of Arbitrarily Shaped Objects*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (1994), pp573-580.
- [SAMP93] Sampsell JB. *An Overview of Texas Instruments Digital Micromirror Device (DMD) and Its Application to Projection Displays*. Society for Information Display Internatl. Symposium Digest of Tech. Papers, Vol. 24, (May 1993), pp1012-1015.
- [SATO97] Sato Y, Wheeler MD, and Ikeuchi K. *Object Shape and Reflectance Modeling from Observation*. Proceedings of ACM SIGGRAPH 97, in Computer Graphics Proceedings, Annual Conference Series 1997, ACM SIGGRAPH, (Aug 1997), pp379-387.
- [SATO99] Sato I, Sato Y, and Ikeuchi K. *Acquiring a radiance distribution to superimpose virtual objects onto a real scene*. IEEE Transactions on Visualization and Computer Graphics, 5(1) (1999), pp1-12.
- [SCHL41] Schlosberg H. *Stereoscopic depth from single pictures*. American Journal of Psychology, 54, (1941), pp601-605.
- [SEIT97] Seitz C and Dyer C. *Photorealistic scene in reconstruction by voxel coloring* In International Conference on Computer Vision and Pattern Recognition, (1997), pp1067-1073.
- [SEQU99] Sequeira V, Ng K, Wolfart E, Gonçalves JGM and Hogg D. *Automated reconstruction of 3D Models from Real Environments Photogrammetry and Remote Sensing*, 54(1), (Feb 1999), pp1-22.
- [SHAU90] Shaughnessy J and Zechmeister E. *Research Methods in Psychology*, (2nd edition), McGraw-Hill, (1990).
- [SIEG84] Seigel R and Howell JR. *Thermal Radiation and Heat Transfer*. Washington DC, Hemisphere Publishing, (1984).
- [SLAB00] Slabaugh G, Malzbender T and Culbertson WB. *Volumetric warping for voxel coloring on an infinite domain*. In M. Pollefeys, L. V. Gool, A. Zisserman, and A. Fitzgibbon, 3D Structure from Images - SMILE 2000.
- [SMIT95] Smith AR., *Image Compositing Fundamentals*, Tech Memo 4, Microsoft, (Jun 1995).

- [SONY03] (www.sony.com)
- [SPSS03] SPSS Inc. Headquarters, 233 S, Wacker Drive, Chicago, Illinois, USA, (2003).
- [STRA88] Strang G. *Linear Algebra and Its Applications*. Harcourt, Brace and Jovanovich, (1988).
- [SUTH68] Sutherland IE. *A head-mounted three-dimensional display*. In Proc. the Fall Joint Computer Conference, (1968), pp757-764.
- [TATH97] Tatham EW. *Depth Cueing for Augmented Reality*. Proceedings of Int. Conf. on Information Visualization, IV'97, London, (August 1997), pp348-349.
- [TATH99a] Tatham EW. *Optical Occlusion and Shadows in a See-through Augmented Reality Display*. Proceedings of Int. Conf. on Information Visualization, IV'99, London, (Aug 1997), pp128-131.
- [TATH99b] Tatham EW. *Getting the Best of Both Real and Virtual Worlds*. Communications of the ACM, Vol. 42, No.9, (Sept 1999), pp96-98.
- [TATH99c] Tatham EW. *Inferring a Plausible Diffuse Illumination Model in an Unbounded Environment*. Proceedings of 6<sup>th</sup> UKVRSIG Conference, Salford, UK, (Sept 1999), pp177-185.
- [TOMI00] Tominaga S and Tanaka N. *Estimating Reflection Parameters from a Single Color Image*. IEEE Computer Graphics and Applications 20 (2000), pp58-66.
- [TSAI87] Tsai RY. *A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using off-the-shelf TV Cameras and Lenses* IEEE Journal of Robotics and Automation, RA-3(4), (Aug 1987), pp323-344.
- [UCHI02] Uchida T. Graduate School of Engineering Science, Dept of Systems Innovation Division of Systems Science and Applied Informatics, Osaka University, Japan, 2002.
- [VLAH58] Vlahos P. *Composite Photography Utilizing Sodium Vapor Illumination*. U. S. Patent 3,095,304, (May 15, 1958). Expired.
- [VLAH64] Vlahos P. *Composite Color Photography, The classic color-difference blue screen compositing technique*. US Patent 3,158,477, (Nov 24, 1964).
- [WALK94] Walker M. *Ghostmasters* Cool Hand Communications, 1994 (ISBN 1-56790-146-8)



[WALL89] Wallace JR, Elmquist KA, and Haines EA, *A Ray Tracing Algorithm for Progressive Radiosity*, Computer Graphics 23, No. 3, SIGGRAPH '89 Proceedings, (Aug 1989), pp315-324.

[WATT89] Watt A. *Fundamentals of Three-Dimensional Computer Graphics*. Addison-Wesley Publishing, (1989).

[WATZ99] Watzinger F, Birkfellner W, Wanschitz F, Millesi W, Schopper C, Sinko K, Huber K, Bergmann H, and Ewers R. *Positioning of dental implants using computer-aided navigation and an optical tracking system: Case report and presentation of a new method*, in J. Craniomaxfac. Surg. 27(2), (1999), pp77-81.

[WEBS96] Webster A, Feiner S, MacIntyre B, Massie W and Krueger T. *Augmented Reality in Architectural Construction, Inspection, and Renovation*, in Proceedings of Computing in Civil Engineering, ASCE, (1996), pp91.

[WELC99] Welch G, Bishop G, Vicci L, Brumback S, Keller K and Colucci D. *The HiBall Tracker: High-Performance Wide-Area Tracking for Virtual and Augmented Environments*. Proceedings of the ACM Symposium on Virtual Reality Software and Technology 1999 (VRST 99), University College London, (Dec 1999), pp1-10.

[WHIT80] Whitted T. *An Improved Illumination Model for Shaded Display* Communications of the ACM 23, No. 6, (June 1980), pp343-349.

[WISM78] Wismer DA and Chattergy R. *Introduction to Nonlinear Optimization*, Elsevier-North Holland, (1978).

[WLOK95] Wloka M, and Anderson B. *Resolving occlusion in augmented reality*. ACM SIGGRAPH 1995 Symposium on Interactive (3D) Graphics, Hanrahan P. and Winget J. (eds), (1995), pp5-12.

[YONA78] Yonas A, Goldsmith LT, and Hallstrom JL. *Development of sensitivity to information provided by cast shadows in pictures*. Perception. 7, (1978), pp333-341.

[YOU99] You S, Neumann U, Azuma R. *Hybrid Inertial and Vision Tracking for Augmented Reality Registration*. Proceedings of IEEE Virtual Reality 1999 (VR'99), IEEE Computer Society Press, Los Alamitos, CA, (1999), pp. 260-267.

[YU99] Yu Y, Debevec ., Malik J, and Hawkins T. *Inverse Global Illumination: Recovering Reflectance Models of Real Scenes from Photographs*. SIGGRAPH, (1999), pp215-224.

[ZEME88] Zemeckis R. (director) *Who Framed Roger Rabbit*. Touchstone Pictures/Amblin Entertainment, Buena Vista, Robert Zemeckis, (1988).

[ZHAN94] Zhang R, Tsai P-S, Cryer JE and Shah M. *Analysis of Shape from Shading Techniques*, IEEE CVPR-94, Washington, (June 1994), pp377-384, 1994.

[ZHAN99] Zhang D and Herbert M. *Harmonic Maps and their Application in Surface Matching* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (1999), pp524-530.



## Appendix A - Camera calibration

### A.1 Objective

To calibrate ray tracer view angle to that of Canon Powershot 600 Digital Camera.

### A.2 Theoretical basis

Using the digital camera, and a set camera to object distance, the correspondence between object width in millimetres and image width in pixels can be determined.

Camera image size is inversely proportional to object distance, so;

$$\text{image width in pixels} = k / (\text{distance in mm}) \text{ for a particular object width} \\ \text{and where } k \text{ is the proportionality constant.}$$

Thus the distance at which the image size in pixels numerically matches the object size in millimetres can be determined. For any given object size, the distance to object width ratio required to produce a 'same-size' (millimetre equivalent to pixel) image can then be calculated. Setting the ray tracer with this distance-to-object ratio, then measuring the resultant image widths for different horizontal view angles, allows the ray tracer view angle, that gives perspective to match the camera, to be determined.

### A.3 Method/Results

A ruler was photographed, laying flat in the plane 1000mm from the camera lens and parallel to the camera image plane. The resultant image was scaled to screen size 640x480 pixels. Correspondence between real world and final digitised image pixels were determined to be:

**In the horizontal direction:**

509mm in real world was found to produce a 475 pixel width image at distance of 1000mm on screen image 640x480 pixels.

**In the vertical direction:**

509mm in real world was also found to produce a 475 pixel width image at distance of 1000mm on screen image 640x480 pixels.

As:

$$\text{image width in pixels} = k / (\text{distance in mm}) \text{ for a particular object width} \\ \text{and where } k \text{ is the proportionality constant.}$$

then:

$$k = (\text{image width in pixels}) \times (\text{distance in mm})$$

Therefore:

$$k = 475 \times 1000 = 475000 \text{ (pixel.mm) for } 509\text{mm object}$$

Thus, a 509mm object would give a 509 pixel image at:

$$\text{distance} = k / \text{image width in pixels} = 475000 / 509 = 933\text{mm (approx.)}$$

i.e. the camera produces 1 pixel / mm (in the horizontal and vertical directions) at about 933mm object distance for 640x480 pixel final image scaling.

As the camera produces 1 pixel / mm at about 933mm, it would give 311 pixels / 311mm at the same object distance.

That is, the camera gives 311 pixel width when:

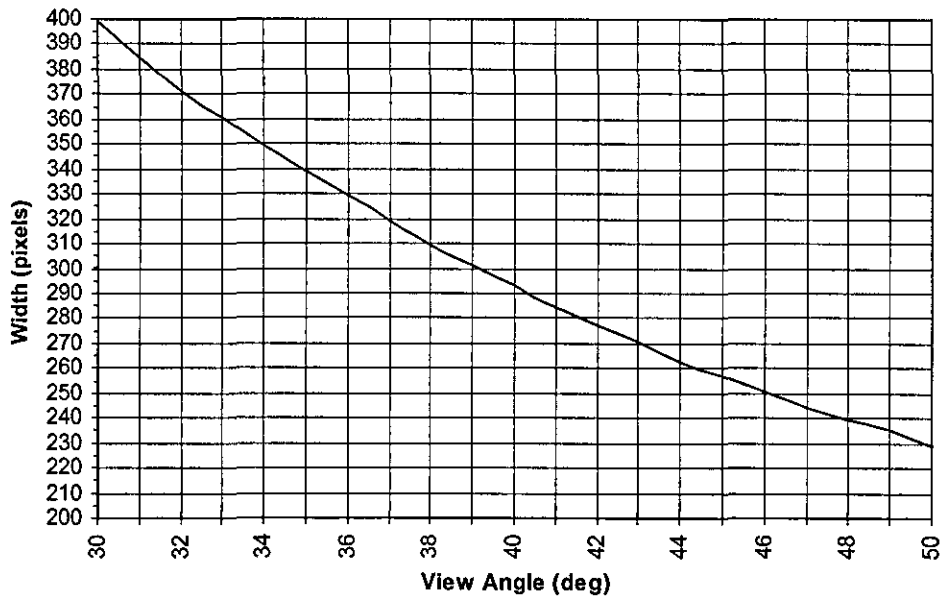
$$\text{width ratio} = 933 / 311 = 3$$

Adjusting the ray tracer view angle for a 3:1 distance: object width ratio gave the following image widths on a 640x480 image.

| View Angle (deg) | Width (pixels) | View Angle (deg) | Width (pixels) |
|------------------|----------------|------------------|----------------|
| 30               | 399            | 41               | 284            |
| 31               | 385            | 42               | 277            |
| 32               | 371            | 43               | 271            |
| 33               | 361            | 44               | 263            |
| 34               | 349            | 45               | 257            |
| 35               | 339            | 46               | 251            |
| 36               | 329            | 47               | 244            |
| 37               | 319            | 48               | 239            |
| 38               | 309            | 49               | 235            |
| 39               | 301            | 50               | 229            |
| 40               | 293            |                  |                |



Change in image width with View Angle  
for 3:1 dist:obj size ratio



From the graph it can be determined that a 311 pixel image is produced, for a 3:1 distance:object width ratio, when the horizontal view angle is set to 37.9°.

## Appendix B - Instructions to experiment participants

**Dear Participant,**

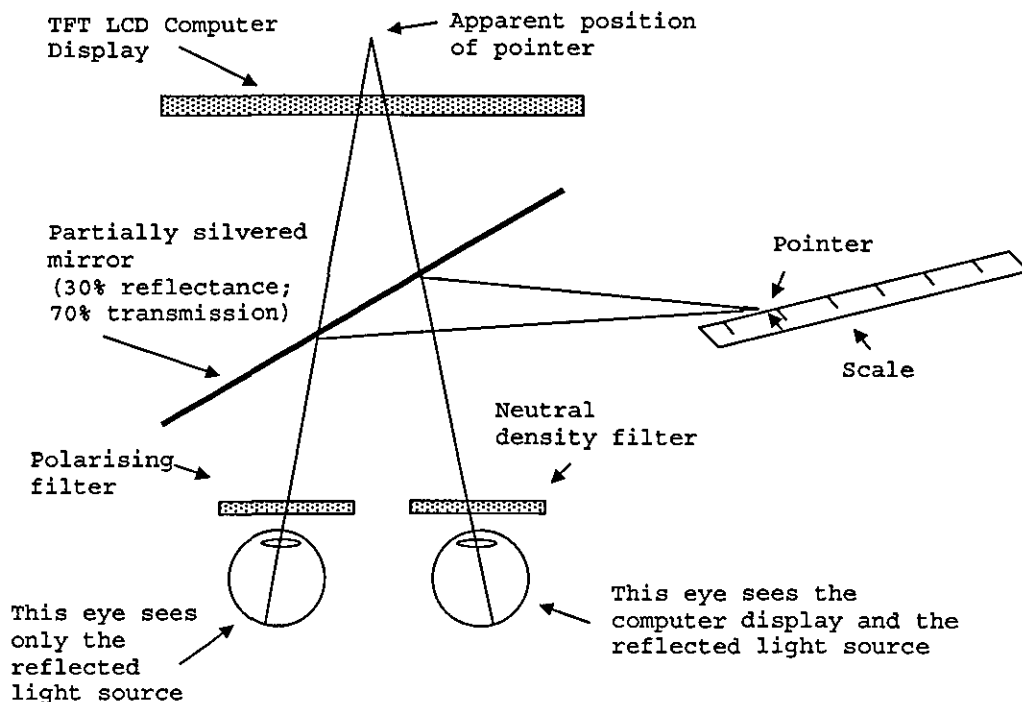
Thank you for agreeing to help with this experiment. It will require about 30 minutes of your time.

The purpose is to measure the effectiveness of a variety of depth cues in computer systems that combine computer graphics with images of real scenes.

In order that the experiment generates useful results, you need to be aged between 18 and 24, have had no need at any time to wear spectacles or contact lenses, and have no known eye defects.

Before participating in the main experiment you will be asked to take a quick test to access your ability to judge distances.

The main experiment is based on the apparatus described below.



The LCD computer display is naturally polarised. This is viewed with one eye through a neutral density filter and a part-silvered mirror angled at  $45^\circ$  with respect to the screen. Placed in front of the other eye is a polarising filter orientated at right-angles to the first polarising sheet so that this eye cannot see the computer screen. A small pointer is arranged so both eyes can see its reflection in the partially silvered mirror in such a way that it appears to come from within the displayed scene. Thus, the display image is viewed with one eye whilst the pointer is viewed with both eyes. As the pointer is moved backward and forward, its reflection should appear at varying depth within the displayed scene. In principle, the pointer can thus be adjusted until its apparent depth within the scene coincides with the object whose depth is to be determined.

You will be asked to view 10 images, through this apparatus, and to gauge the depth of a specified object by asking the experimenter to move a pointer backward and forward until it appears to be at the same depth as the object.

**Appendix C - Depth experiment result sheets**

**Participant Number:** 1

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 11 | Image 0 near pillar              | 11 |
| Image 0 far pillar              | 32 | Image 0 far pillar               | 35 |
| Image 3                         | 28 | Image 3                          | 32 |
| Image 4                         | 17 | Image 4                          | 18 |
| Image 2                         | 10 | Image 2                          | 10 |
| Image 1                         | 10 | Image 1                          | 9  |

**Participant Number:** 2

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 13 | Image 0 near pillar              | 14 |
| Image 0 far pillar              | 26 | Image 0 far pillar               | 31 |
| Image 4                         | 21 | Image 4                          | 26 |
| Image 1                         | 13 | Image 1                          | 13 |
| Image 3                         | 15 | Image 3                          | 20 |
| Image 2                         | 12 | Image 2                          | 11 |

**Participant Number:** 3

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 10 | Image 0 near pillar              | 13 |
| Image 0 far pillar              | 27 | Image 0 far pillar               | 28 |
| Image 1                         | 11 | Image 1                          | 12 |
| Image 2                         | 10 | Image 2                          | 10 |
| Image 4                         | 19 | Image 4                          | 20 |
| Image 3                         | 11 | Image 3                          | 17 |

**Participant Number:** 4

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 11 | Image 0 near pillar              | 10 |
| Image 0 far pillar              | 33 | Image 0 far pillar               | 29 |
| Image 2                         | 8  | Image 2                          | 10 |
| Image 3                         | 18 | Image 3                          | 14 |
| Image 1                         | 9  | Image 1                          | 9  |
| Image 4                         | 22 | Image 4                          | 20 |

Participant Number: 5

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 9  | Image 0 near pillar              | 11 |
| Image 0 far pillar              | 25 | Image 0 far pillar               | 31 |
| Image 3                         | 17 | Image 3                          | 23 |
| Image 4                         | 21 | Image 4                          | 24 |
| Image 2                         | 8  | Image 2                          | 10 |
| Image 1                         | 8  | Image 1                          | 10 |

Participant Number: 6

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 12 | Image 0 near pillar              | 12 |
| Image 0 far pillar              | 30 | Image 0 far pillar               | 32 |
| Image 4                         | 26 | Image 4                          | 27 |
| Image 1                         | 11 | Image 1                          | 10 |
| Image 3                         | 16 | Image 3                          | 15 |
| Image 2                         | 11 | Image 2                          | 10 |

Participant Number: 7

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 13 | Image 0 near pillar              | 11 |
| Image 0 far pillar              | 35 | Image 0 far pillar               | 29 |
| Image 1                         | 11 | Image 1                          | 12 |
| Image 2                         | 12 | Image 2                          | 10 |
| Image 4                         | 26 | Image 4                          | 24 |
| Image 3                         | 21 | Image 3                          | 19 |

Participant Number: 8

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 12 | Image 0 near pillar              | 11 |
| Image 0 far pillar              | 28 | Image 0 far pillar               | 26 |
| Image 2                         | 8  | Image 2                          | 7  |
| Image 3                         | 18 | Image 3                          | 13 |
| Image 1                         | 10 | Image 1                          | 9  |
| Image 4                         | 19 | Image 4                          | 20 |



**Participant Number:** 9

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 10 | Image 0 near pillar              | 12 |
| Image 0 far pillar              | 31 | Image 0 far pillar               | 35 |
| Image 3                         | 18 | Image 3                          | 22 |
| Image 4                         | 26 | Image 4                          | 27 |
| Image 2                         | 9  | Image 2                          | 11 |
| Image 1                         | 8  | Image 1                          | 9  |

**Participant Number:** 10

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 14 | Image 0 near pillar              | 13 |
| Image 0 far pillar              | 34 | Image 0 far pillar               | 36 |
| Image 4                         | 22 | Image 4                          | 25 |
| Image 1                         | 12 | Image 1                          | 12 |
| Image 3                         | 26 | Image 3                          | 28 |
| Image 2                         | 12 | Image 2                          | 12 |

**Participant Number:** 11

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 12 | Image 0 near pillar              | 12 |
| Image 0 far pillar              | 29 | Image 0 far pillar               | 27 |
| Image 1                         | 11 | Image 1                          | 9  |
| Image 2                         | 9  | Image 2                          | 8  |
| Image 4                         | 26 | Image 4                          | 23 |
| Image 3                         | 18 | Image 3                          | 16 |

**Participant Number:** 12

| <i>Left eye cross-polarised</i> |    | <i>Right eye cross-polarised</i> |    |
|---------------------------------|----|----------------------------------|----|
| Image 0 near pillar             | 11 | Image 0 near pillar              | 10 |
| Image 0 far pillar              | 26 | Image 0 far pillar               | 29 |
| Image 2                         | 9  | Image 2                          | 7  |
| Image 3                         | 19 | Image 3                          | 23 |
| Image 1                         | 10 | Image 1                          | 9  |
| Image 4                         | 17 | Image 4                          | 19 |

## Appendix D – Provisional Patent Application

### NEW BRITISH PATENT APPLICATION

Type: Provisional Patent Application

Applicant(s): Coventry University

Title: Computer augmented reality display

Reference:

Application no: 9807107.9

Filing Date: 03 APR 98

Lewis & Taylor  
5 The Quadrant  
Coventry CV1 2EL



The  
Patent  
Office

FILING RECEIPT

EN Lewis & Taylor (Coventry)  
5 The Quadrant  
Coventry  
West Midlands  
UK  
CV1 2EL

The Patent Office

Concept House  
Cardiff Road  
Newport  
South Wales  
NP9 1RH

Switchboard  
01633-814000

Your Ref. : P5376UK

03 April 1998

PATENT APPLICATION NUMBER      9807107.9

The Patent Office confirms receipt of a request for grant of a patent, details of which have been recorded as follows :

|                                      |                       |
|--------------------------------------|-----------------------|
| Filing Date (See Note)               | : 03-APR-98           |
| Applicants                           | : Coventry University |
| Description (No. of Sheets)          | : 4                   |
| Claims (No. of Sheets)               | : None                |
| Drawings (No. of Sheets)             | : 2+2                 |
| Abstract                             | : None                |
| Statement of Invention (Form 7/77)   | : None                |
| Request for Search (Form 9/77)       | : None                |
| Request for Examination (Form 10/77) | : None                |
| Priority Documents                   | : None                |
| Translation of Priority Documents    | : None                |
| Divisional of Application            | : None                |
| Divisional Date Claimed              | :                     |
| Other Attachments Received           | : None                |

The application number included in the heading above should be quoted on all correspondence with The Patent Office.

Any queries on this receipt should be addressed to Mrs Lynne Payne, tel 01633 814570.

Note : The above filing date is provisional and may need to be amended if the provisions of section 15(1) of the Patents Act 1977 are not met.

Patents Form 1/77

Patents Act 1977  
(Rule 16)



**Request for grant of a patent**

*(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)*

The Patent Office

Cardiff Road  
Newport  
Gwent NP9 1RH

1. Your reference P5376UK

2. Patent application number  
*(The Patent Office will fill in this part)*

3. Full name, address and postcode of the or of each applicant. *(underline all surnames)* Coventry University  
Priory Street, Coventry, CV1 5FB

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation Great Britain

4. Title of the invention Computer-augmented reality display

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom Lewis & Taylor, 5 The Quadrant, Coventry, CV1 2EL  
to which all correspondence should be sent  
*(including the postcode)*

Patents ADP number *(if you know it)* 711001

| 6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and <i>(if you know it)</i> the or each application number | Country | Priority application number <i>(if you know it)</i> | Date of filing <i>(day / month / year)</i> |
|---|---------|---|--|
|   |         |   |  |

| 7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application | Number of earlier application | Date of filing <i>(day / month / year)</i> |
|---|-------------------------------|--|
|   |                               |  |

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) *any applicant named in part 3 is not an inventor, or*

b) *there is an inventor who is not named as an applicant, or*

c) *any named applicant is a corporate body*

*See note (d))* Yes

**Patents Form 1/77**

9. Enter the number of sheets for any of the following items you are filing with this form.  
Do not count copies of the same document

Continuation sheets of this form 0  
Description 4  
Claim(s) 0  
Abstract 0  
Drawing(s) 2

10. If you are also filing any of the following, state how many against each item.

Priority documents  
Translations of priority documents  
Statement of inventorship and right to grant of a patent (Patents Form 7/77)  
Request for preliminary examination and search (Patents Form 9/77)  
Request for substantive examination (Patents Form 10/77)  
Any other documents 0  
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature *Lewis Taylor* Date *2/4/98*

12. Name and daytime telephone number of A V Hallam  
person to contact in the United Kingdom 01203 222756/227145

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication of communication of the invention should be prohibited or restricted under Section 22 or the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

**Notes**

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patent Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.



COV5176.LJK

1

### Computer- augmented reality display

The present invention relates to a computer-augmented reality display.

Virtual reality computer systems place a participant in a virtual environment in which as many human senses as possible are isolated from real world experience and are fed by  
5 computer stimuli. However, a fully immersive virtual reality approach suffers from a number of drawbacks.

It is usually desirable for the virtual environment to be as realistic as possible in order to give it the credibility that will allow the user to suspend disbelief. However, artificially  
10 simulating anything that approximates a real world environment requires considerable computing power.

Rather than replacing the real environment with a wholly artificial environment, hitherto known systems have used computers to augment reality. In one such system a see-through head-mounted display is employed enabling the user to see the real environment through  
15 part-silvered mirrors that also reflect a visually superimposed graphic image into the user's eyes. Another system uses a conventional virtual reality headset to provide a non-see-through augmented reality display in which the user sees a video image of reality combined with luminance or chroma-keyed graphics. However, real-time visual intergration of graphics with the real world presents significant difficulties. There is a need, for example, to ensure that, where appropriate, real objects appear in front of virtual, occluding parts that  
20 cannot be seen and to appear as if lit by the light sources that exist in reality, and vice-versa. Existing augmented reality display strategies fundamentally preclude effective integration for systems where a direct non-video view of reality is required or preferred. The transparent overlay nature of current see-through displays allows occlusion of virtual objects and can produce some shadowing effects. However, with such systems it is not possible to interpose  
25 opaque virtual objects in front of the real. On the other hand, the straightforward luminance or chrominance keying used in non-see-through systems is suited to opaque composition but

COV3276 UK

2

cannot achieve the necessary transparent overlay effects.

The present invention seeks to provide an improved computer-augmented reality display system.

Accordingly, the present invention provides a computer-augmented reality display system  
5 comprising:

visual control means for positioning in a line of sight between a viewer and a real scene;

and computer means for generating a graphic display;

wherein said visual control means comprises:

10 a beam splitter for reflecting light from said graphic image to the viewer thereby enabling said viewer to view said graphic image superimposed on said real scene;

and an active panel means controlled by said computer for selectively controlling passage of light from said real scene towards said viewer.

The present invention is further described hereinafter, by way of example, with reference to the accompanying drawings in which:

15 Figure 1 is an example of a known display arrangement for augmenting visual reality; and

Figure 2 is a preferred form of display arrangement according to the present invention for augmenting visual reality with computer-generated images.

Referring firstly to figure 1, this shows a known display arrangement which is in the form

COV5174 LK

3

of a see-through head-mounted display 10. The viewer 12 looks at the real scene 14 through a beam splitter 16 which is at an angle of 45 degrees to the line of sight 15 between the viewer 12 and the real scene.

5 A graphic display 18 is generated by a computer laterally of the beam splitter such that the display is reflected by the beam splitter back along the line of sight 15 towards the viewer 12.

This known system has the disadvantages previously set out above.

Referring now to figure 2, this shows a preferred form of display arrangement 20 according to the present invention for augmenting visual reality with computer generated images.

10 The system comprises a beam splitter 22 which is conveniently in the form of a part or half-silvered mirror. This is generally planar and lies at typically 45 degrees to the line of sight 24 of the viewer 26. The mirror 22 is positioned such that the viewer 26 can see the real scene 28 through the mirror.

15 A display 29 which may conveniently be an LCD display is positioned laterally of the mirror 22 such that light from the display 29 is incident on the mirror at right angles to the line of sight 24 and is reflected by the mirror along the line of sight to the viewer 26. Images are generated on the LCD display by a computer with the arrangement being such that the viewer 26 can see the computer generated images as if they were superimposed on the real scene 28.

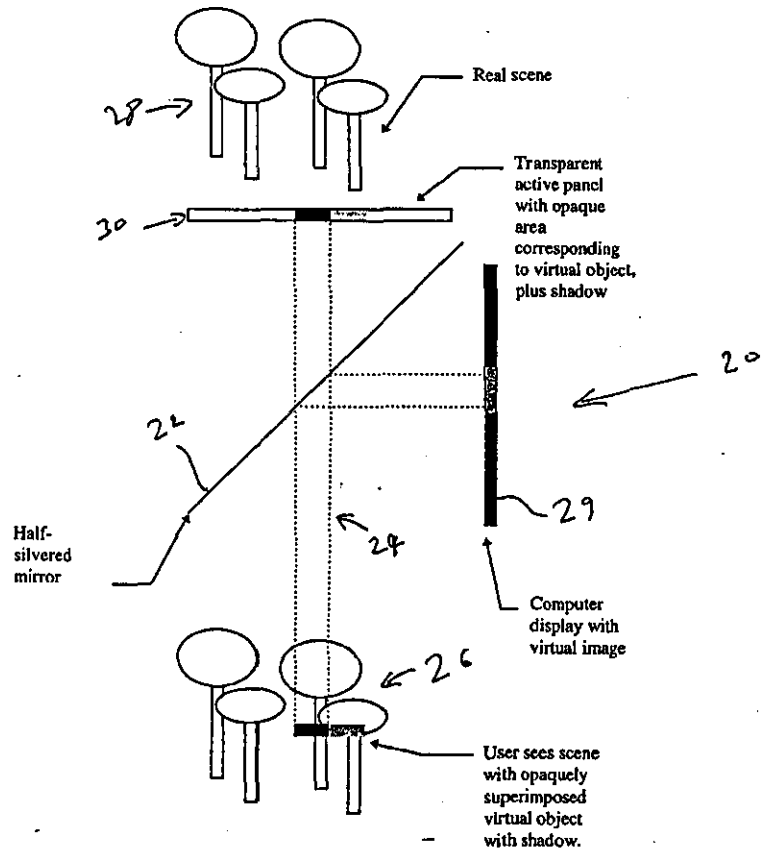
20 In order to provide occlusion and shadowing effects, a transparent active panel 30 is positioned between the real scene 28 and the mirror 22. The active panel is capable of displaying a solid mask corresponding to opaque areas of the virtual objects generated in the display screen 28, and can produce density or colour variations to create shadow and other visual effects which are impossible with existing display arrangements.

COV3176 UK

4

The transparent active filter 30 may conveniently be an LCD screen although other suitable screens may be used.

- The screen 30 is conveniently controlled by the computer which generates the image on the screen 28 (it may be a different computer but then both computers would need to be synchronised) such that the active screen 30 is capable of displaying a solid mask corresponding to opaque areas of the virtual objects and can produce density or colour variations to create shadow and other visual effects impossible with existing display arrangements. It is also possible to create other real-virtual visual interaction such as colour bleeding and highlights on real objects created by virtual light sources.
- 10 It will be appreciated that it is possible to combine the functions of the beam splitter 22 and the active screen 30 in a single device or panel, or in a coating or layer on a screen or panel.





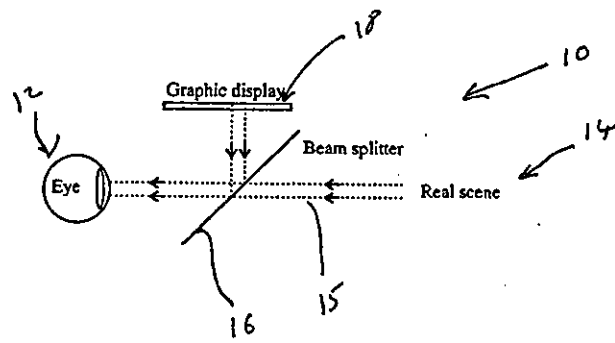


Fig 1.

## Appendix E – Papers based on work described in the thesis

This appendix contains copies of the following papers:

[TATH97] Tatham EW. *Depth Cueing for Augmented Reality*. Proceedings of Int. Conf. on Information Visualization, IV'97, London, (August 1997), pp348-349.

[KALA98] Kalawsky RS and Tatham EW. *Effects of Spatial and Temporal Mis-registration in Augmented Virtual Environments*, Proceedings of the 16th Annual Conference, Eurographics UK, Leeds, (March 1998), pp127-134.

[TATH99a] Tatham EW. *Optical Occlusion and Shadows in a See-through Augmented Reality Display*. Proceedings of Int. Conf. on Information Visualization, IV'99, London, (Aug 1997), pp128-131.

[TATH99b] Tatham EW. *Getting the Best of Both Real and Virtual Worlds*. Communications of the ACM, Vol. 42, No.9, (Sept 1999), pp96-98.

[TATH99c] Tatham EW. *Inferring a Plausible Diffuse Illumination Model in an Unbounded Environment*. Proceedings of 6<sup>th</sup> UKVRSIG Conference, Salford, UK, (Sept 1999), pp177-185.

## Depth Cueing for Augmented Reality

Eric W Tatham  
 School of Mathematical and Information Sciences  
 Coventry University  
 UK

### Abstract

*Computer-augmented reality systems promise to overcome some of the problems inherent in virtual reality and, in addition, to provide numerous other application possibilities. However, if augmented reality is to become fully practicable a number of hurdles need to be surmounted. Not least of these is the ability to achieve convincing real-time visual integration of virtual objects within real scenes. To this end, it is particularly important that virtual entities appear to exist at their appropriate depth within a real environment. This poster outlines an experimental system to determine the efficacy of alternative static monocular depth cues in an augmented reality display.*

### Introduction

Unlike immersive virtual reality systems that seek to place the human operator in a wholly artificial environment, augmented reality systems aim to supplement views of reality through the super-imposition of computer graphics. There are two basic ways in which this can be achieved. A see-through head-mounted display can be employed enabling the user to see the real environment through part-silvered mirrors that also reflect a visually superimposed graphic image into the user's eyes or, alternatively, a conventional virtual reality head-set can be used to provide a non-see-through augmented reality display in which the user sees a video image of reality combined with luminance or chroma-keyed graphics. For some applications it is sufficient to dispense with a head-mounted display and to present the video-graphic image using a conventional monitor, viewed either monoscopically, or stereoscopically with shutter glasses. Also some form of tracking system is usually required to detect the viewing position, such that the synthetic imagery can be appropriately registered with the real.

Augmented reality systems promise a large number of potential applications spanning a diversity of fields, from manufacturing and maintenance (e.g. Caudell and Mizell,

[1], or Feiner, et al., [2]), to medical imaging and design visualisation (e.g. Bajara, et al., [3], Gunkel et al., [4], and Ahlers et al., [5]). However, before this promise can be fulfilled, a number of practical problems must be resolved.

### Seamless integration

Accurate registration of graphics and reality presents a significant problem due to the limited range and accuracy of current tracking devices. A further related problem stems from the latency evident between a change of view of the real world and the corresponding computer graphic update. Additionally, for some envisaged applications, there is a need for even more convincing integration of synthetic and real imagery than can be achieved through accurate registration alone. Seamless synthesis must depend on a number of factors relating to the way in which virtual objects appear and visually interact with a real scene. For example, virtual objects placed in a real environment would be expected to be occluded by real objects that are nearer the viewer (Breen et al., [6]; Wloka and Anderson, [7]), and to appear as if lit by the light sources that exist in reality. Similarly,



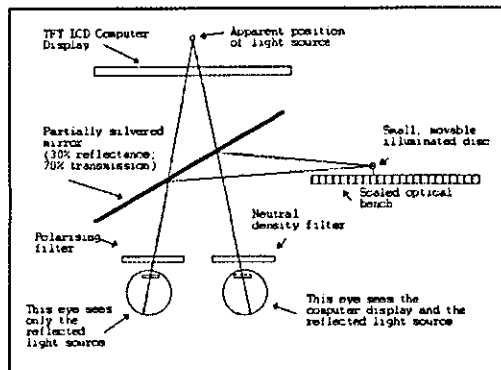
A virtual box reflected in a real mirror

glossy or mirrored surfaces in reality should reflect appropriately placed virtual objects, and the real should be reflected in the virtual.

For successful integration, it is of fundamental importance that virtual entities are perceived as being correctly sited within the environment depth. There are a number of so-called cues that can account, at least in part, for the raw data our brains require for effective depth perception. Some of these cues rely on binocular vision, some on relative motion, whilst others are able to provide information about depth to even a stationary, monocular viewer. A further complication, in the case of computer displays, is that depth cues may be countered by 'flatness cues', that mitigate against the perception of depth.

### The experimental system

To determine the relative efficacy of static monocular depth cues in an augmented display, an experiment has been designed and apparatus constructed based in principle on that employed some years ago by Deregowski [8] in relation to work on measuring the apparent depth of objects in drawings. The system has been built cheaply using plastic piping to form a closed optical system that, as far as possible, reduces the flatness cues normally apparent in the computer display.



An image of the real scene with superimposed computer generated graphics is displayed on a flat TFT LCD screen. The screen is viewed with one eye through both a neutral density filter, and a partially silvered mirror angled at 45 degrees with respect to the computer display. Placed in front of the other eye is a polarising filter orientated at right-angles to the screen polariser so that this eye cannot see the

computer image. A small illuminated disc is mounted on an optical bench arranged orthogonally to the line of sight so that both eyes can see its reflection in the partially silvered mirror in such a way that it appears to emanate from within the computer screen. Thus, the screen image is viewed monocularly whilst the reflection of the disc is viewed binocularly. As the disc is moved backward and forward along the optical bench, its reflection appears at varying depth within the displayed scene. The disc can thus be adjusted using binocular vision until its apparent depth within the image coincides with the monocularly displayed object whose apparent depth is to be determined. The actual position of the disc can be determined from the scaled optical bench.

With this equipment we are seeking to quantify the effectiveness of a variety of monocular cues in an augmented reality system.

### References

1. Caudell, T., and Mizell, D. Augmented reality: An application of heads-up display technology to manual manufacturing processes, Proc. Hawaii Int. Conf. System sciences, Jan 1992, 659-669
2. Feiner, S., MacIntyre, B., and Seligmann, D. Annotating the real world with knowledge-based graphics on a see-through head-mounted display, Proc. Graphics Interface 1992, Canadian Info. Proc. Soc, 78-85
3. Bajura, M., Fuchs, H., and Ohbuchi, R. Merging virtual objects with the real world: seeing ultrasound imagery within the patient, Computer Graphics, 26 (2), 1992, 203-210
4. Gunkel, A.R., Freysinger, W., Thumfart, W.F., and Truppe, M.J. Application of the ARTMA image-guided navigation system to endonasal sinus surgery, Proc. (Computer Assisted Radiology) CAR95, Berlin, June 21-24, 1995, 1146-1151
5. Ahlers, K., Kramer, A., Breen, D., Chevalier, P.Y., Crampton, C., Rose, E., Tuceryan, M., Whitaker, R., and Greer, D. Distributed augmented reality for collaborative design applications. Proc. Eurographics '95 Conf., Maastricht, Netherlands, August 1995, 3-14
6. Breen, D., Rose, E., Whitaker, R. Interactive occlusion and collision of real and virtual objects in augmented reality, Technical Report ECRC-95-02, ECRC, Munich, Germany, 1995
7. Wloka, M., Anderson, B. Resolving occlusion in augmented reality, in ACM SIGGRAPH 1995 Symposium on Interactive (3D) Graphics, Hanrahan, P., and Winget, J. (eds), 5-12
8. Deregowski, J. Pictorial Depth and Culture, Scientific American, No. 227, Nov 1972, 82-88

## Effects of spatial and temporal mis-registration in augmented virtual environments

**Roy S. Kalawsky**

Advanced VR Research Centre,  
Loughborough University,  
Loughborough, Leicestershire, LE11 3TU, UK,  
Tel +44 (0)1509 223047, +44 (0)1509 223940,  
Email: r.s.kalawsky@lboro.ac.uk,  
WWW: <http://sgi-hursk.lboro.ac.uk/~avrcc/>

**Eric W. Tatham**

School of Mathematical and Information Sciences,  
Coventry University,  
Coventry, West Midlands, CV1 5FB, UK,  
Tel +44 (0)1203 838909,  
Email: e.w.tatham@coventry.ac.uk,  
WWW: <http://www.mis.coventry.ac.uk/~EricT/>

### Abstract

Computer-augmented reality systems have the potential to overcome some of the problems inherent in virtual reality and, in addition, to provide many new application possibilities. However, if augmented reality is to become fully practicable, a number of hurdles must first be surmounted. Not least of these is the ability to achieve convincing real-time visual integration of virtual objects within real scenes. To this end, it is particularly important that virtual entities register correctly and appear situated at their appropriate depth within their real environment. This paper focuses on these issues, presenting initial results from on-going work aimed at improving spatial integration in computer-augmented realities.

Over the past few years, there has been widespread interest in the development of computer systems that place the human operator in a virtual environment in which as many human senses as possible are isolated from real world experience and fed by computer stimuli. This concept of virtual reality has caught the imagination of a wider public, and research continues to devise techniques to increase realism, through faster and more convincing graphics, as well as the provision of effective haptic feedback enabling the sense of touch to be added to the virtual experience. Such systems promise numerous applications in areas such as education, training, entertainment and design visualisation. However, a fully immersive, virtual reality approach suffers from a number of drawbacks.

One problem currently being researched is how to avoid the feelings of vertigo and nausea that can arise when using virtual reality systems. (Biocca, 1992.) Additionally, in many applications it is desirable for the virtual environment to be as realistic as possible to give it the credibility that will allow its users to suspend disbelief. This requirement drives the development of environment models that seek to better represent our real world perceptions. However, artificially simulating anything even loosely approximating a real world environment carries a huge computational overhead. Considerable effort is required to provide significant levels of detail in areas that may not even be the focus of the task, such as realistic landscapes for flight simulators and walk-throughs for architectural designs. For collaborative systems a further overhead may be introduced by the need to model virtual representations of the human participants. A further constraint imposed by immersive virtual reality systems is that users are insulated from the real world and therefore cannot readily interact with it in order to carry out real tasks.

*Kalawsky/Tatham*

Rather than replacing the real environment with one that is wholly artificial, a number of early researchers (e.g. Sutherland, 1968; Furness, 1969; Knowlton, 1977 and Krueger et al., 1985, 1991) have sought to use computers to augment real experience. There are two basic ways in which this can be achieved: firstly, a see-through head-mounted display can be employed enabling the user to see the real environment through part-silvered mirrors that also reflect a visually superimposed graphic image into the user's eyes and secondly, a conventional virtual reality headset can be used to provide a non-see-through augmented reality display in which the user sees a video image of reality combined with luminance or chroma-keyed graphics. (Kalawsky 1991, 1992, 1993.) For some applications it is sufficient to dispense with a head-mounted display and to present the video-graphic image using a conventional monitor or television projection system, viewed either monoscopically or stereoscopically with shutter glasses.

Computer enhancement of reality offers advantages over virtual reality by not only potentially avoiding the need for complex modelling of people and environment, but also by providing an anchor in reality that should reduce the likelihood of nausea being induced. In addition, augmented reality systems promise to allow the operator to actively carry out tasks involving real world objects rather than being confined to a wholly artificial environment. A wide range of potential applications exist, and include - assistance for manufacturing and maintenance (e.g. Caudell and Mizell, 1992 or Feiner et al., 1992), medical imaging (e.g. Bajura et al., 1992 or Gunkel et al., 1995), annotating the real world (e.g. Rose et al., 1995), training (e.g. Metzger, 1993), tele-operation of robots (e.g. Milgram et al., 1995), design visualisation (e.g. Ahlers et al., 1995) and collaborative working (e.g. Rekimoto, 1996).

The development of systems to fulfil such applications is the subject of current research but it is very much in its infancy. Significant problems still need to be resolved if any are to be truly viable. In particular, real-time registration of graphics with the real world presents a significant challenge, and is crucial to the success of most applications, especially those providing manufacturing and surgical guidance. Whereas small tracking inaccuracies may not be noticeable in an immersive virtual reality system, even very small angular errors in detecting the orientation of an augmented reality headset will result in a large displacement in *registration of graphics with real objects that are some distance away*.

The use of head mounted displays to provide information overlays to personnel is a concept that has been used by the military for over two decades. The most important and highly developed systems are used by pilots of military aircraft (fixed and rotary wing). Early generation systems presented symbolic graphical overlays onto the real world. Information included aircraft heading, aircraft attitude, altitude, airspeed and special targeting symbology. This data represents information that is earth referenced, aircraft referenced and pilot stabilised. More recently, information derived from thermal imaging systems and terrain databases has been provided to the pilot by the head mounted display. The benefits are immense, so that when information from the real environment is compromised (e.g. poor weather conditions or night time) it is possible to provide an *augmented/enhanced display*. A critical aspect of the design of these systems is the need to provide accurate registration between the real and virtual environments. Large scale mis-registration errors occur when there are inconsistencies between the real environment and the onboard database, for example, new electricity pylons not being referenced in the database. This is an extreme example but serves to illustrate the point that a mis-registered database can cause serious problems.

Today augmented VR systems are not restricted to head mounted displays. Alternative display technologies are available and include wide angle projection displays (monoscopic and stereoscopic) as well as traditional desktop displays.



Whenever, an image is overlaid onto the real world (either via an optical or electronic system) there are many sources of error including, scaling, translational, angular and temporal. When any of these errors are present the overlaid image will not register correctly with the real world causing particular problems for a user which can lead to an increased workload or hazardous situation. It is important to recognise that there are two types of mis-registration error. These are static and dynamic. The effect of each mis-registration error is different and requires a different solution to minimise its effects. It is important to note that there is no such thing as perfect registration. Instead we must think in terms of registration tolerances.

Static registration refers to scale, translation and angular correspondence between the real and virtual environment representation. Any resulting mis-registration is visible as a mismatch between corresponding points (landmarks) in the real and virtual environments. Units of measurement tend to be scale, translational or angular. The latter two being specified in linear or angular measurements.

Dynamic registration refers to the correspondence between the real and virtual worlds as either (or both) are moved with respect to one another. A common unit of measurement for dynamic registration is to measure the time lag between the real and virtual image. Linear or angular accelerations can also be used to express the mis-match. The visual manifestation of dynamic registration errors are initially hard to spot in all but extreme cases. The typical way it shows up is the virtual object appearing to have a variable or changing position with respect to the real world and being a complex function of the rate of change of movement. In some systems it is possible for the virtual image to appear to catch up with the real world as the user's movements slow down. Even though it is possible to achieve extremely good static accuracy with a tolerance of  $\pm 0.5\text{mm}$  and  $\pm 0.1^\circ$  it is possible to have a poor dynamic accuracy of greater than 10ms lag. In this case everything looks fine for a static observer but as soon as any movement is taking place then the image suffers from unusual lags.

Figure 1 (Kalawsky 1998) shows the primary sources of temporal registration error in a VR system. The arrows illustrate that different parts of the system - the underlying simulation, the graphical generation of the virtual scene and the refresh rate of the display devices each have their own cycle time and hence introduce a small lag into the image finally seen by the user. The various lags are additive and the precise relationship between the individual time lags is a complicated non linear function. Unless special care is taken in the implementation of the graphics system it is highly likely that the update rate will be a function of scene complexity and thus is a variable quantity which fluctuates during use. Many advanced graphics systems support constant frame rates meaning that each frame is updated at a predefined frame rate. Unless a reliable update rate is achieved it will be difficult to develop techniques to reduce the mis-registration.

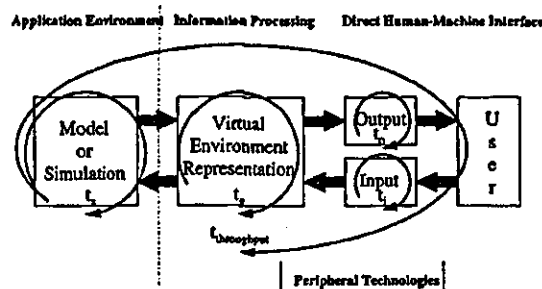


Figure 1: Sources of Dynamic Mis-registration in a VR system

*Kalawsky/Tatham*

The optical system can introduce unintentional distortions into the virtual or real world scene. Without going into the specific design of a particular optical system it is possible for the combining plate and associated optics to introduce field curvature to the virtual environment. The effect tends to become worse the further one gets away from the optic axis of the optical system.

A further complication arises whenever a stereoscopic image is produced by a display system. Firstly, any mis-match between the display system's inter-ocular distance (IPD) and the user's IPD leads to an apparent change in perceived depth. Generally, as the display system's IPD is increased with respect to the user the stereoscopic effect is exaggerated. This in turn leads to a translational scaling error in the depth axis. Unless a careful calibration process is followed it is very difficult to spot this type of mis-registration error. A related error occurs when a camera based system is used to create an augmented display. The camera must be at the same point in space as the user's eye otherwise an offset will be introduced. The displacement error causes potential problems in user interaction.

The greatest difficulty comes in measuring the actual static and dynamic mis-registration. The process used in the Advanced VR Research Centre is to use an adapted surveying technique. At the moment it is a time consuming process necessitating a large number of angular measurements. Time sequential stereoscopic modes present a different challenge and are outside the scope of this paper. Understanding the user's task is important because it will enable the impact of any mis-registration to be understood. The way a user will interact with the system is of great importance as other side effects may be present. The presence of a moving mis-registered image could lead to the onset of simulator sickness because the brain is trying to compensate for the errors. In an extreme case the virtual display can look as though it is swimming around. It has even been known for the real world to appear as though it is unstable and results in a most unpleasant feeling.

Once the magnitude of the mis-registration error has been determined there are a range of techniques available to compensate for the effect. The most satisfactory solutions involve optical correction (ideally through a re-design of the optical system) because this does not introduce a computational overhead. Static registration errors can be easily compensated by adding known offsets into the graphics system to compensate for the constant error. Error correction algorithms for dynamic registration can be developed but unless these are executed on a fast processor it is possible for the correction algorithm to lead to increased dynamic registration problems. Various techniques have been employed to try to predict the extent of the temporal lag and provide some form of lead compensation. If these solutions are inappropriately designed then further problems can occur. Work is ongoing to develop appropriate solutions to the mis-registration problem.

Mis-registration effects do not only manifest themselves in a visual sense. It is also possible for the user interaction devices such as 3D joysticks or mice to be incorrectly calibrated. This could lead to the user having to make inappropriately scaled movements in the virtual environment. This none 1:1 correspondence with the real world means that the user has to adapt or compensate for the differences. At times this can be used to advantage where the user effectively scales their movements to enable large distances to be traversed without having to move very far. Generally, this makes the interface difficult to use as the user has to re-adapt to the new environment. For some applications, the overlaid image complexity may be low, with only simple graphics for which mis-registration is hardly noticed. However, the success of more sophisticated augmented reality systems may ultimately depend on the seamlessness with which synthetic graphics and reality can be merged. Poorly merged graphics do not have the required credibility and can cause attentional difficulties. Stokes et al. (1990), in a review of the literature on head-up displays, noted the problem associated with

operators having to divide attention between real-world and synthetic imagery. This difficulty may be particularly acute in applications that rely on the ability of the system to allow suspension of disbelief by convincingly integrating the virtual with the real. Seamless synthesis will depend on a number of other factors, besides registration, where virtual objects would be expected to visually interact with a real environment, or vice-versa. There is a need, therefore, to consider factors relating to the physical and visual interaction between real and virtual entities. It is of fundamental importance that, where appropriate, real objects appear in front of virtual, occluding parts that cannot be seen (Breen et al., 1995; Wloka and Anderson, 1995) and to appear as if lit by the light sources that exist in reality. Similarly, glossy or mirrored surfaces in reality should reflect appropriately placed virtual objects, and the real should be reflected in the virtual. Real shadows may fall across virtual objects and virtual shadows across real objects. Virtual light sources would be expected to disperse real shadows and vice-versa. Reality should appear refracted through transparent graphics, and virtual objects refracted by the real. Atmospheric effects such as fog, smoke, heat haze, or just plain air should affect the appearance of virtual objects in the same way as they do reality.

Some of these interactions seem reasonably straightforward whilst others appear inherently intractable. Nevertheless, for successful integration, it is of fundamental importance that virtual entities are perceived as being sited correctly within the environment depth, and that we have ways of assessing the verity of this perception.

Assessing the apparent depth of objects that are, in reality, on a flat display screen presents a challenge. However, we have built an apparatus for the purpose of determining the efficacy of monocular depth cues, based on an original design used by R.Gregory (1977) to view the Muller-Lyer illusion and thus provide evidence in support of his size constancy theory.

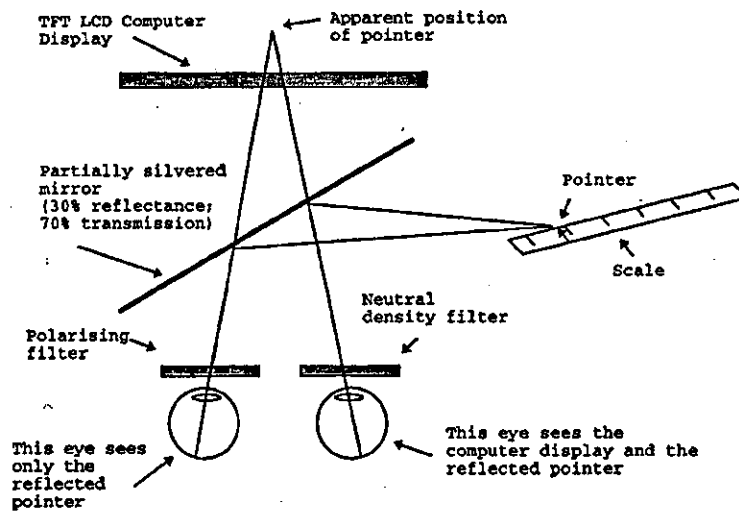


Figure 2: Adaptation of Gregory's Virtual Depth Apparatus

*Kalawsky/Tatham*

An image of the real scene with superimposed computer generated graphics is displayed on a flat TFT LCD screen. The screen is viewed with one eye through both a neutral density filter, and a partially silvered mirror angled at 45 degrees with respect to the computer display. Placed in front of the other eye is a polarising filter orientated at right-angles to the screen polariser so that this eye cannot see the computer image. An illuminated pointer is mounted on a scale arranged to appear in the part-silvered mirror, superimposed on the scene so that both eyes can see its reflection. Thus the screen image is viewed monocularly whilst the reflection of the pointer is viewed binocularly. As the pointer is moved backward and forward along the scale, its reflection appears at varying depth within the displayed scene. The pointer can thus be adjusted using binocular vision until its apparent depth within the image coincides with the monocularly displayed object whose apparent depth is to be determined. The screen is viewed monocularly in order to reduce its apparent flatness. To this end, the viewing is through a circular tube that removes the flatness cue afforded by a rectangular frame.

We have utilised this apparatus in an experiment that verifies and quantifies the importance of realising appropriate occlusion and shadow effects if effective visual integration is to be achieved in augmented reality systems. The results highlight a predictable problem with existing augmented reality display strategies that fundamentally preclude effective integration. The transparent overlay nature of current see-through displays allows occlusion of virtual objects and can produce some shadowing effects. However, with such systems, it is not possible to interpose opaque virtual objects in front of the real. On the other hand, the straightforward luminance or chrominance keying used in non-see-through systems is suited to opaque composition but cannot achieve the necessary transparent overlay effects.

It seems likely that some of the issues relating to smooth integration of the real and the virtual will prove difficult to resolve. However, towards this end, we have begun work exploring alternative display strategies to overcome these current limitations.

Augmented virtual reality promises to be extremely important in the future because it provides a way of exploiting real world cues which are difficult to emulate in a computer system. What is certain though, is the user's eye-brain's capability to resolve very fine detail and be very discriminating when scene information is conflicting with other sensory cues. On the other hand the eye-brain is very forgiving and can tolerate certain ambiguities without causing any problems for the user. The goal is to overcome or exploit these effects to allow us to produce better content for virtual environments. The human factors basis for our research is intended to ensure that we understand the complicated underlying science.

## References

- Ahlers, K., Kramer, A., Breen, D., Chevalier, P.Y., Crampton, C., Rose, E., Tuceryan, M. Whitaker, R., and Greer, D. (1995) Distributed augmented reality for collaborative design applications. Proc. Eurographics '95 Conf., Maastricht, Netherlands, August 1995, pp3-14.
- Bajura, M., Fuchs, H., and Ohbuchi, R. (1992) Merging virtual objects with the real world: seeing ultrasound imagery within the patient, *Computer Graphics*, 26 (2), pp203-210.
- Biocca, F. (1992) Will simulation sickness slow down the diffusion of virtual reality technology? Working Paper Series: Communication Technology and Cognition Group, Centre for research in Journalism and Mass Communication, University of North Carolina.

*Kalawsky/Tatham*

Breen, D., Rose, E., Whitaker, R. (1995) Interactive occlusion and collision of real and virtual objects in augmented reality, Technical Report ECRC-95-02, ECRC, Munich, Germany.

Caudell, T., and Mizell, D. (1992) Augmented reality: An application of heads-up display technology to manual manufacturing processes, Proc. Hawaii Int. Conf. System sciences, Jan 1992, pp659-669.

Feiner, S., MacIntyre, B., and Seligmann, D. (1992) Annotating the real world with knowledge-based graphics on a see-through head-mounted display, Proc. Graphics Interface 1992, Canadian Info. Proc. Soc, pp78-85.

Furness, T. (1969) Helmet-Mounted Displays and their Aerospace Applications, National Aerospace Electronics Conference, Dayton OH, May 1969.

Gregory, R. L. (1977) Eye and Brain - The Psychology of Seeing, Weidenfeld and Nicolson, 1977.

Gunkel, A.R., Freysinger, W., Thumfart, W.F., and Truppe, M.J. (1995) Application of the ARTMA image-guided navigation system to endonasal sinus surgery, Proc. (Computer Assisted Radiology) CAR95, June 21-24, Berlin, 1146-1151.

Kalawsky R.S. (1991) "Integrated Real-Virtual World Display System" Patent Specification No. 9121707.5, 12 Oct

Kalawsky R.S., (1992) "Integrated real-world virtual world system", Patent Specification US 07/959,919 24 Apr

Kalawsky R.S., (1992) "Computer Generated Images with Real View Overlay" European Patent Specification 92309158.1

Kalawsky R.S., (1993) "The Science of Virtual Reality and Virtual Environments" Addison Wesley

Kalawsky, R.S. (1998) *VRUSE* A Computerised Diagnostic Tool: For Usability Evaluation of Virtual/Synthetic Environment Systems, Applied Ergonomics, in press

Knowlton, K. (1977) Computer displays optically superimposed on input devices, The Bell Syst, Tech. Journal, 56, 3, March 1977, 367-383.

Krueger, M., Hinrichsen, K., Gionfriddo, T. (1985) VIDEOPLACE - An artificial reality, Proc. SIGCHI 1985, 35-40.

Krueger, M. (1991) Artificial Reality II, Addison-Wesley, Reading, Mass.

Metzger, P.J. (1993) Adding reality to the virtual, Proc. IEEE Virtual Reality Annual Int. Symp., September 18-22, Seattle, WA, 7-13.

Milgram, P., Drascic, D., and Grodski, J., Restogi, A., Zhai, S., and Zhou, C. (1995) Merging real and virtual worlds, Proc of IMAGINA '95, Monte Carlo, Feb 1-3, 1995.

*Kalawsky/Tatham*

Rekimoto, Jun (1996) TransVision: A hand-held augmented reality system for collaborative design in Virtual Systems and Multi-Media (VSMM) '96.

Rose, E., Breen, D., Ahlers, K., Crampton, C., Tuceryan, M., Whitaker, R., and Greer, D. (1995) Annotating real-world objects using augmented reality, Proc. of Computer Graphics: Developments in Virtual Environments International 95, June 1995.

Stokes, A., Wickens, C. and Kite, K. (1990) Display Technology: Human Factors Concepts, Warrendale, PA: Society of Automotive Engineers, Inc.

Sutherland, I. (1968) A head-mounted three dimensional display, Proc. FJCC 1968, Thompson Books, Washington DC, 1968, 757-764.

Wloka, M., Anderson, B. (1995) Resolving occlusion in augmented reality, in ACM SIGGRAPH 1995 Symposium on Interactive (3D) Graphics, Hanrahan, P., and Winget, J. (eds), 5-12.



## Optical Occlusion and Shadows in a 'See-through' Augmented Reality Display

Eric W Tatham  
Augmented Reality Research Group  
Faculty of Mathematics and Computing  
The Open University  
Walton Hall  
Milton Keynes, MK7 6AA  
UK  
Tel. +44 (0)1908 655098  
Fax. +44 (0)1908 652140  
Email e.w.tatham@open.ac.uk

### Abstract

*As distinct from virtual reality, which seeks to immerse the user in a fully synthetic world, computer-augmented reality systems supplement sensory input with computer-generated information. The principle has, for a number of years, been employed in the head-up display systems used by military pilots and usually comprises an optical display arrangement based on part-silvered mirrors that reflect computer graphics into the eye in such a way that they appear superimposed on the real-world view. Compositing real and virtual worlds offers many new and exciting possibilities but also presents some significant challenges, particularly with respect to applications for which the real and virtual elements need to be integrated convincingly. Unfortunately, the inherent difficulties are compounded further in situations where a direct, unpixelated view of the real world is desired, since current optical systems do not allow real-virtual occlusion, nor a number of other essential visual interactions. This paper presents a generic model of augmented reality as a context for discussion, and then describes a simple but effective technique for providing a significant degree of control over the visual compositing of real and virtual worlds.*

Superimposing electronic graphics on our view of the real world is a familiar feature of SLR (Single Lens Reflex) cameras that typically present exposure and other information superimposed over the

photographer's view through the lens. It is evident that there is further potential in using such an arrangement for providing visual information adaptable to given situations and, indeed, this was exploited by Knowlton [7] in 1977 when he developed a system that visually superimposed computer displays onto an input device. The purpose was to allow users to interact with a real physical keyboard whilst also providing flexibility of function by optically superimposing alternative labels onto the keys. In this way, the same physical keyboard could be endowed with the appearance of a typewriter, a calculator or a telephone operator's console.

The essential generic components of such a computer-augmented reality system are illustrated in figure 1. Perceptual stimuli from the real environment are augmented by computer generated elements to provide a composite perceptual experience. Depending on the system's purpose, it is normally necessary for the synthetic elements to be harmonised in some way with the real. Usually, this will require that the synthesising computer have access to information about pertinent aspects of the world, such as; world geometry, user position and orientation, illumination, or physical object and atmospheric properties. For simplicity the model shows the augmentation system as external to the user's environment. Although this generic model is intended to be applicable for all kinds of other stimuli, such as augmentation of a real environment with computer-generated music, it is primarily real-virtual visual integration on which many new applications depend.

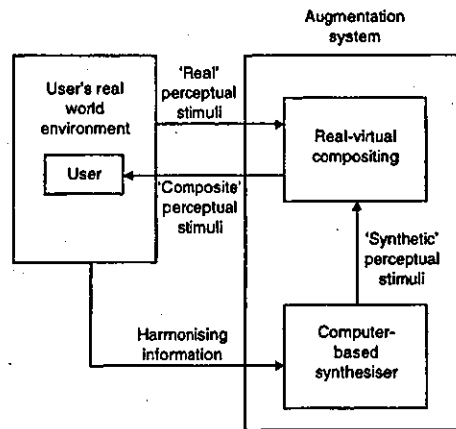


Figure 1

A wide range of possibilities exist for usefully augmenting reality and includes; assistance for manufacturing and maintenance [4,5], medical imaging [3], annotating the real world [11], training [8], collaborative working [10]. However, a number of teleoperation of robots [9], design visualization [1] and technical problems need to be resolved if the promise of augmented reality is to find fruition. Almost all potential applications depend on the acquisition and utilisation of appropriate harmonising information, and this often presents a significant hurdle. For example, accurate spatial and temporal registration of computer graphics onto a real scene remains crucial to the success of most applications, especially those providing manufacturing or surgical guidance. Whereas small tracking inaccuracies may not be noticeable in an immersive virtual reality system, even very small angular errors in detecting the orientation of an augmented reality headset can result in a large displacement in the registration of the graphics. Ultimately, more accurate methods of position and orientation tracking are required, as well as effective methods of tracking over larger distances. Further registration problems can occur due to latency between changes in the real scene and the corresponding computer graphics update, as the graphics almost inevitably lag behind. The total delay is caused by the period it takes for the tracker subsystem to take its measurements and the time for the corresponding images to appear on the display devices. Not surprisingly, these problems provide the focus for much of the current research effort [2], but seamless visual integration of real and virtual worlds also depends on

effective simulation of other factors [6]. It is of fundamental importance in conveying a convincing perception of depth that, where appropriate, real objects appear in front of virtual, occluding parts that cannot be seen, and that virtual objects be suitably interposed before real. Achieving such interpositioning is obviously dependent on the availability of knowledge concerning the depth of real-world objects. However, occlusion is just the tip of a much larger iceberg of interactions that need to be resolved if convincing integration is to be realised. For example, virtual objects placed in a real environment should be expected to appear as if lit by the light sources that exist in reality. In turn, but less tractable, is the case where the virtual object is itself a light source that would be expected to illuminate the real environment. In practice, almost all objects will reflect some light that will add to the illumination of nearby objects; thus some of the colour from a red virtual object would appear to bleed into a real matt white surface on which it is placed, and vice-versa. Glossy or mirrored surfaces in reality should reflect appropriately placed virtual objects, and the real should be reflected in the virtual. Real shadows may fall across virtual objects and virtual shadows across real objects. Reality should appear refracted through transparent graphics, and virtual objects refracted by the real. In addition, atmospheric effects such as fog, smoke or heat haze should affect the appearance of virtual objects in the same way as they do real.

There is little doubt that acquisition and utilisation of appropriate harmonising information present the most pressing challenges for augmented reality

researchers. However, consideration of the required visual interaction effects exposes an inherent limitation associated with current augmented reality display hardware in its ability to achieve effective real-virtual compositing. In Knowlton's system the compositing was achieved by reflecting the computer graphics in a semi-transparent mirror that was positioned over the keyboard at a strategic angle, and it is this same principle that is employed, although greater freedom of movement obtained, by current see-through, head-mounted displays; see figure 2.

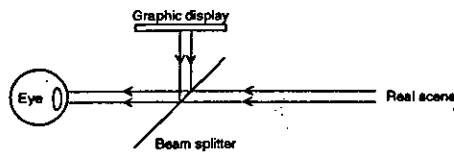


Figure 2

A similar visual effect, but at the expense of stereopsis, is obtained by presenting the computer graphics to just one eye whilst leaving the other eye free to view the real scene directly so that the task of superimposition of the two components is left for the brain to complete. Unfortunately, both arrangements produce an overlaid image that is transparent and ghost-like, providing no control over light from the real world and precluding the possibility of occluding real objects by the virtual. An alternative compositing technique that overcomes this particular problem, and is often used, is to key computer graphic elements into a video image of the real scene. However, such an arrangement requires head-mounted cameras, pixellates the real-world view and, in the event of hardware failure, obscures the user's vision.

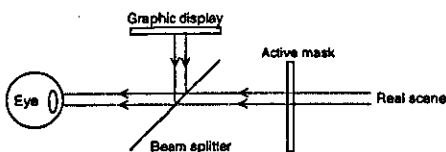


Figure 3

In order to retain the direct-view advantages of a see-through augmented reality display while also providing occlusion and other desired visual interaction effects, a modified hardware set-up is proposed. The basic alteration to existing displays is very simple but effective and is illustrated in figure 3. It is based on the introduction of an active filter capable of masking portions of the real scene. In the illustrated

arrangement, a computer-generated image is viewed by reflection in a part-silvered mirror as before. However, the real world is now viewed, not only through the mirror, but also via a transparent active panel that is placed along the viewer's line of sight. The computer display and transparent panel image are both spatially registered and temporally synchronised, with the transparent element acting as an active mask selectively reducing the intensity of light from the world that reaches the viewer's eye. On the other hand, the reflected image selectively increases the light reaching the eye. This arrangement allows for significant flexibility and control as it is now possible to both reduce and increase the intensity and colour of light reaching the viewer from selected areas of a scene.

Using the transparent display element to create an opaque mask and the reflected element to display the superimposed graphic object allows virtual entities to visually occlude a real background. The active mask can also be used to generate areas of neutral density that reduce light received from selected portions of the real world enabling the simulation of virtual shadows within a real scene. The left-hand image in figure 4, showing a shaded sphere with black background, is reflected into the eye by the beam splitter (see figure 3) while the shadow image with transparent background, shown on the right in the figure, is displayed as the active mask. These images combine to form one of the frames for an animated sequence in which the virtual sphere orbits a real Lego™ pillar. The sphere occludes its real background and appears to cast its shadow on reality as it moves. Figure 5 shows a single frame from the video sequence. Figure 6 is a photograph taken directly through the display to show a real finger inserted through a virtual torus. Besides producing occlusion and shadow effects, using an active colour mask colour permits selective filtering of areas of real world colour. Thus, employing such a mask facilitates production of any desired colour bleeding or illumination effects.

Although the illustrations in this article have been produced, for convenience, using a LCD panel as the active mask, this is not ideal since such panels introduce limiting attenuation and distortion. Fortunately, there are possible alternative methods of realising the required active masking and these form the subject of the author's current research effort. Promising designs that dispense entirely with the need for an active transparent panel, as well as obviating the requirement for a part-silvered mirror, are likely to be based on the use of spatial light modulator devices. Whatever the hardware used for implementation, the principle of incorporating active masking, as outlined in this paper, overcomes many of the inherent limitations of current see-through displays. Such a technique provides a degree of control that could help enable computer-augmented reality to fulfil its promise of becoming a highly versatile tool for the future.

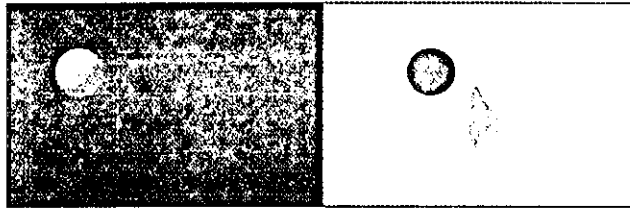


Figure 4



Figure 5

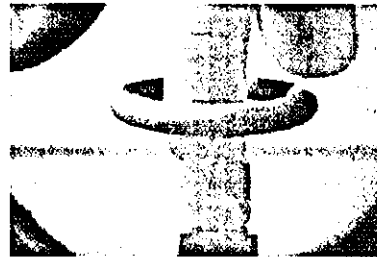


Figure 6

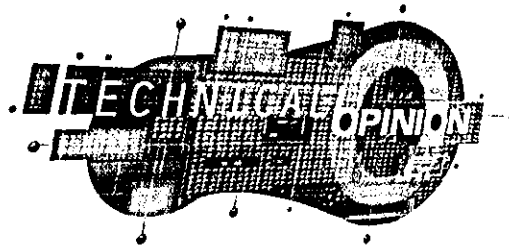
#### Acknowledgements

The ideas described in this paper arose as a result of work supported by the author's former employer; Coventry University, UK. Thanks are also due to R. Kalawsky of Loughborough University and N.Godwin, K. Monk and R.Newman all of Coventry University.

#### References

1. Ahlers, K., Kramer, A., Breen, D., Chevalier, P.Y., Crampton, C., Rose, E., Tuceryan, M., Whitaker, R., and Greer, D. Distributed Augmented Reality for Collaborative Design Applications. Proc. Eurographics '95 Conf., (Maastricht, Netherlands, August 1995), 3-14.
2. Azuma, R. Tracking Requirements for Augmented Reality. Communications of the ACM, Vol. 36 (7), (July 1993), 50-51.
3. Bajura, M., Fuchs, H., and Ohbuchi, R. Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient, Computer Graphics, 26 (2), (1992), 203-210.
4. Caudell, T., and Mizell, D. Augmented reality: An Application of Heads-up Display Technology to Manual Manufacturing Processes, Proc. Hawaii Int. Conf. System Sciences, (January 1992), 659-669.
5. Feiner, S., MacIntyre, B., and Seligmann, D. Annotating the Real World with Knowledge-Based Graphics on a See-through Head-mounted Display, Proc. Graphics Interface 1992, Canadian Info. Proc. Soc. (1992), 78-85.
6. Kalawsky, R.S. and Tatham, E.W. Effects of Spatial and Temporal Mis-registration in Augmented Virtual Environments, Proc. of the 16<sup>th</sup> Annual Conference of Eurographics, UK, Leeds, (March 1998), 127-134.
7. Knowlton, K. Computer Displays Optically Superimposed on Input Devices, The Bell Syst, Tech. Journal, 56 (3), (March 1977), 367-383.
8. Metzger, P.J. Adding Reality to the Virtual, Proc. IEEE Virtual Reality Annual Int. Symp., (Seattle, WA, September 1993), 7-13.
9. Milgram, P., Drascic, D., Grodski, J., Restogi, A., Zhai, S., and Zhou, C. Merging Real and Virtual Worlds, Proc of IMAGINA '95, (Monte Carlo, February 1995), 218-230.
10. Rekimoto, Jun. TransVision: A Hand-held Augmented Reality System for Collaborative Design. Sony Computer Science Laboratory Inc., Takanawa Muse Building, 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo 141, Japan, (1996).
11. Rose, E., Breen, D., Ahlers, K., Crampton, C., Tuceryan, M., Whitaker, R., and Greer, D. Annotating Real-world Objects using Augmented Reality, Proc. of Computer Graphics: Developments in Virtual Environments International 95, (June 1995), 357-370.

Footnote: Lego™ is a trademark of the Lego Group.



*Eric W. Tatham*

## Getting the Best of Both Real and Virtual Worlds

*A simple but effective way to facilitate visual occlusion using a see-through display.*

In contrast to virtual reality systems, generally designed to immerse the user as fully as possible within a synthetic environment, computer-augmented reality supplements real-world stimuli with computer-generated elements. Visually, this is achieved by electronic or optical superimposition of computer graphics with a user's view of the real world. The potential applications are wide ranging, but there are a number of hurdles to overcome if such systems are to reach fruition.

The greatest challenges relate to maintenance of accurate spatial and temporal registration of real and virtual entities when objects are moved or the point of view changes. However, even if these problems are resolved, inherent limitations associated with current "see-through" displays remain when it comes to producing convincing integration of real and virtual elements. Most problematic, they produce a graphic overlay that is transparent and easily washed-out in regions where the background is

bright and, hence, are incapable of simulating the occlusion effects essential for appropriate depth perception. Proposed here is a simple but effective way of facilitating visual occlusion and improving the degree of color control in see-through augmented reality displays.

While most people are familiar with the concept of combining computer-generated graphics with real-world imagery (exemplified, for example, by films such as "Jurassic Park"), achieving similar effects in a head-mounted display over a real-time view of our actual environment is a more difficult proposition. The principle of such computer augmentation of reality has its roots in the early head-mounted display devised by Sutherland [8] and the subsequent head-up displays designed for military pilots. It is only recently that more wide-spread interest and potential has developed.

Visual augmentation offers many of the advantages of a synthetic world while retaining the obvious value of interaction with reality. Getting the best of both

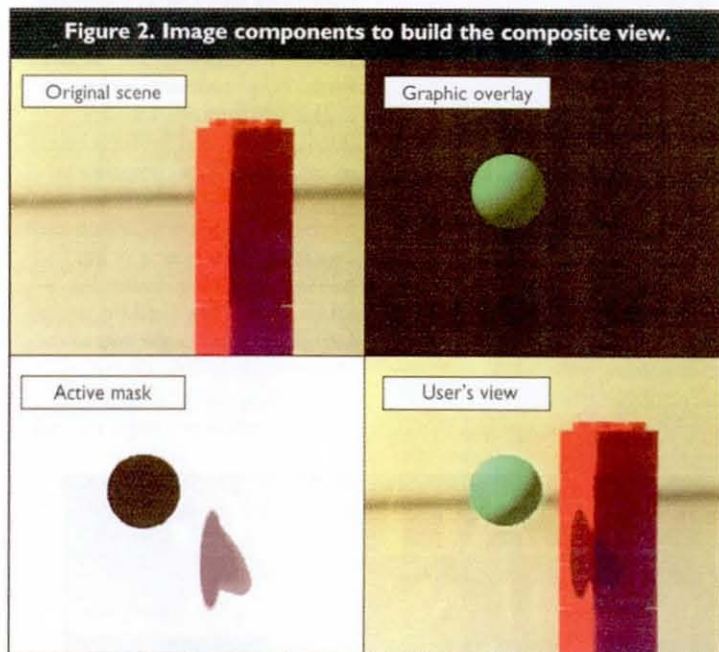
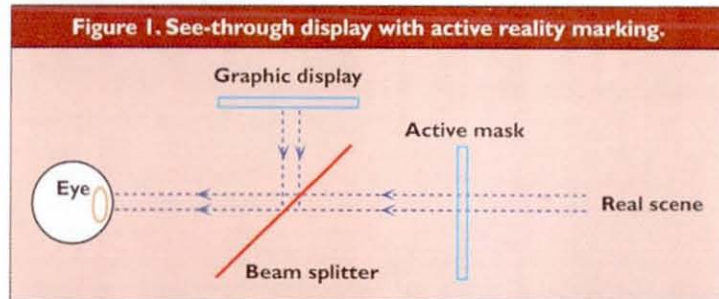
worlds in this way offers many new and exciting possibilities, including assistance in manufacturing and maintenance [4, 5], medical imaging [3], robotic teleoperation [7], and design visualization [1]. However, significant difficulties are inherent, particularly with respect to applications for which the real and virtual elements need to be convincingly integrated. For example, accurate spatial and temporal registration of computer graphics with a real scene remains crucial to the success of most applications, especially those providing manufacturing or surgical guidance. Whereas small tracking inaccuracies might not be noticeable in an immersive VR system, even very small angular errors in detecting the orientation of an augmented reality headset can result in a large displacement in the registration of the graphics.

Ultimately, more accurate methods of position and orientation tracking are required, as well as effective methods of tracking over larger distances. Further registration problems can occur due to

latency between changes in the real scene and the corresponding computer graphic update, as the graphics almost inevitably lag behind. These problems provide the focus for much of the current research in this field [2], but seamless visual integration of real and virtual worlds also depends on effective simulation of other visual factors.

It is of fundamental importance in conveying a convincing depth perception that, where appropriate, real objects appear in front of the virtual, occluding parts that cannot be seen, and virtual objects that are suitably interposed before the real backgrounds. In addition, achieving an accurate composite also depends on realistic simulation of shadows, color-bleeding, and other illumination effects arising from the interaction between real and virtual elements [6].

Even when registration problems are resolved and real-time depth mapping becomes readily available, effective visualization of an augmented composite will not be possible with current see-through displays. These are generally designed so computer-generated images are reflected over the real-world view using partly silvered mirrors angled appropriately in front of each eye. Consequently, the resultant overlay is inherently transparent, ghost-like, and unconvincing. Occlusions cannot be properly represented and the localized adjustment of real-world color quality necessary for simulating changes in illumination is impossible. Although sufficient control is available in augmented reality systems based on a video composite, see-through arrangements are necessary in cases where the synthetic



elements are to be overlaid on an unpixelated view of reality.

In order to retain the direct-view advantages of a see-through, augmented reality display while also providing occlusion and facility for other desired visual interactions, a modified display arrangement is proposed. The basic alteration to existing displays is the addition of an active mask (shown in Figure 1). The computer-generated component is viewed by reflection in a partly silvered mirror. But now the real world is viewed, not only through

the mirror, but also via a transparent active panel placed along the viewer's line of sight.

The computer display and transparent panel image are both spatially registered and temporally synchronized, with the transparent element acting as an active mask, selectively reducing the intensity of light reaching the viewer's eye. On the other hand, the reflected image selectively increases the light reaching the eye. This arrangement allows for significant flexibility and control, making it possible to both reduce and increase the intensity



of light reaching the viewer from any region of a scene. Using the transparent display element to create an opaque mask and the reflected element to display the superimposed graphic object allows virtual entities to visually occlude a real background. The active mask also can be used to generate areas of neutral density that reduce light received from selected portions of the actual world and enable the simulation of virtual shadows within a real scene.

For demonstration purposes, a short animation has been produced in which a virtual sphere appears to orbit a real LEGO pillar (Figure 2). The original scene is shown together with the reflected graphic sphere overlay, its corresponding mask, and composite view seen by the user. Besides producing occlusion and shadow effects, using an active mask capable of filtering color, permits selective filtering of areas of real-world

color. Thus, employing such a mask facilitates production of any desired color-bleeding or illumination effects.

Although the images in Figure 2 were produced using a LCD panel as the active mask, this is far from ideal. Such panels introduce attenuation and distortion. A promising alternative arrangement dispenses entirely with the need for an active transparent panel, as well as obviating the requirement for a partly silvered mirror, through the use of a spatial light modulator such as the Digital Micromirror Device (DMD) developed by Texas Instruments. This device comprises an array superstructure of micro-mechanical aluminum mirrors each associated with a memory bit. Each mirror is an approximately  $16\mu\text{m}$  square and can be tilted electrostatically depending on the state of the underlying memory cell. An appropriate arrangement of optical elements is needed to project the real

and virtual images onto the DMD surface and then onward to the user's eye. Individual mirrors in the DMD can be directed to project portions of a real or virtual image while rapid switching allows the real and virtual to be mixed in varying proportions.

Whatever the hardware used for implementation, the principle of incorporating active reality masking overcomes significant inherent limitations of current see-through displays and provides a degree of control that helps enable computer-augmented reality to fulfill its promise as a highly versatile tool for the future.  $\square$

ERIC W. TATHAM (E.W.Tatham@open.ac.uk) is a Computer Science lecturer at The Open University, U.K.

The ideas described here are from work on spatial integration in computer-augmented reality supported by a Coventry University grant.

## REFERENCES

- Ahlers, K., et al. Distributed augmented reality for collaborative design applications. In *Proceedings of Eurographics '95* (Maastricht, Netherlands, Aug. 1995), 3-14.
- Azuma, R. Tracking requirements for augmented reality. *Commun. ACM* 36, 7 (Jul. 1993), 50-51.
- Bajura, M., Fuchs, H., and Ohbuchi, R. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Comput. Graph.* 26, 2 (1992), 203-210.
- Caudell, T. and Mizell, D. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of Hawaii Int. Conf. System Sciences* (Jan. 1992), 659-669.
- Feiner, S., MacIntyre, B., and Seligmann, D. Knowledge-based augmented reality. *Commun. ACM* 36, 7, (Jul. 1993), 52-62.
- Fournier, A., Gunawan, A., and Romanzin, C. Common illumination between real and computer generated scenes. In *Proceedings of Graphics Interface '93*. (Toronto, May 1993), 254-262.
- Milgram, P., Drascic, D., Grodski, J., Restogi, A., Zhai, S., and Zhou, C. Merging Real and Virtual Worlds. In *Proceedings of IMAGINA '95* (Monte Carlo, Feb. 1995), 218-230.
- Sutherland, I. A head-mounted three Dimensional Display. In *Proceedings of the Fall Joint Computer Conference* (Washington DC, 1968), 757-764.

© 1999 ACM 0002-0782/99/0900 \$5.00

## **Inferring a Plausible Diffuse Illumination Model in an Unbounded Environment**

Eric W. Tatham  
Faculty of Mathematics and Computing  
The Open University, UK

### **Abstract**

Computer-augmented reality systems seek to merge synthetic imagery with views of the real world. However, despite the promise of numerous applications, there are few fully usable systems. Technical problems relating to spatial and temporal registration form a major stumbling block and, for some applications, the convincing visual integration of real and virtual entities presents further significant hurdles. An important prerequisite for seamless integration is the acquisition of a common illumination model but, unfortunately, determining real-world lighting models from the shading of surfaces within the field of view is substantially an intractable problem. This paper presents a strategy for inferring a plausible radiation enclosure to account for diffuse illumination of visible surfaces in a scene where knowledge of illuminants and other surfaces affecting the scene is limited to those within the field of view. The model acquired can be used to re-illuminate the scene after augmentation with new entities. A simple illustrative example is presented.

### **1 Background**

The concept of supplementing views of the real world with computer-generated elements is familiar to anyone who has used the type of single-lens reflex camera that superimposes exposure and other data onto the scene. In a similar way, head-up displays have, for many years, been used for graphically augmenting the pilots' view from military aircraft [7]. The potential for providing augmented visual information adaptable to given situations was exploited further by Knowlton [9] when he developed a system that optically superimposed computer displays onto an input device. Knowlton's purpose was to allow users to interact with a real physical keyboard, whilst also providing flexibility of function by visually superimposing alternative labels onto the keys, giving the same physical keyboard the appearance of a typewriter, a calculator or a telephone operator's console. More recently, a wide range of other application possibilities are being explored, including; assistance for manufacturing and maintenance [3], medical imaging [2], annotating the real world [5][14], training, [11], teleoperation of robots [12], design visualisation [1] and collaborative working [13].

However, a number of technical problems need to be resolved if the promise of augmented reality is to find fruition. For example, accurate spatial and temporal registration of computer graphics onto a real scene remains crucial to the success of most applications, especially those providing manufacturing and surgical guidance. Whereas small tracking inaccuracies may not be noticeable in an immersive virtual reality system, even very small angular errors in detecting the orientation of an augmented reality headset can result in a large displacement in registration of graphics with real objects that are some distance away. Ultimately, more accurate methods of position and orientation tracking are required, as well as effective methods of tracking over larger distances. There are also further registration problems that can occur due to the latency between change of view of the real scene and the corresponding computer graphic update, as the graphics inevitably lag behind changes in the real world. The total delay is caused by the time it takes for the tracker subsystem to take its measurements and the time it takes the corresponding images to appear on the display devices. Not surprisingly, these problems provide the focus for most current augmented reality research effort.

While convincing integration of real and virtual scenes obviously depends on achieving close spatial and temporal registration of the real and virtual parts, it also relies, in significant measure, on harmonisation of the illumination affecting real and virtual objects. A virtual entity introduced into a real scene will alter the balance of light within the scene in ways that can be difficult to determine. For example, there may be real objects that are not initially in view but which affect the illumination of the scene. We would also expect virtual light sources to dissipate real shadows and this requires reinstatement of real-world colour and textures that might not be known. In addition, there may be interactions between hidden parts of real and virtual entities, such as the reflection of the back of a real object in a virtual mirror, or shadows caused by hidden parts of a real object illuminated by a virtual light source. The situation is particularly complex as light radiation bounces from surface to surface within an environment until an equilibrium is reached, effectively making every point on every surface a potential source of reflected, if not emitted, energy that will potentially influence every other point in the environment. Introducing new objects upsets this balance necessitating a global reconfiguration of light interactions. For successful integration, we need to be able to derive an illumination model from a real scene and then apply it to re-illuminate a real-virtual composite. Only if we assume the availability of complete models of the real and virtual worlds, including all light sources, does this proposition appear fully tractable.

With respect to acquiring a common illumination model, Fournier, Gunawan and Romanzin [6] address some of the issues in their work on compositing computer graphics and video images. Their strategy involves modelling objects in the real scene with bounding boxes and then using the video image intensity to deduce the initial surface radiosity of the visible parts of the boxes. Surface reflectances of the boxes are approximated using an estimate of illuminant intensity based on the concept of ambient light, before a progressive radiosity computation and ray-casting are used to render the scene. In this work, viewing parameters are reconciled by a manual process of interactively matching the computer graphics with the video scene, and the environment used is an enclosed room in which all light source positions and sizes are modelled beforehand. Where there is no model of the real light sources, every element is considered to be an emitter. However, the technique does assume that there are no unknown radiative inputs to the system, and that all significant geometry is known. Unfortunately, in many real-world situations, the environment will be unbounded and knowledge of its geometry and light sources will be incomplete. Obtaining a

common illumination model then becomes more difficult as we need to infer the position and intensity of light sources and unseen reflective surfaces that will account for the illumination and consequent shading visible on known surfaces.

For some scenes where sufficient object geometry and surface properties are known or can be determined, in principle, it is possible to track backwards along imaginary rays, from the location of highlights and cast shadows, to provide an indication of the directions to the original light sources. However, such ray-casting techniques cannot deal effectively with light sources that have finite area and, in reality, few sources fulfil the qualification of being point sources. Indeed, the real world is illuminated predominantly with light that is reflected diffusely from surface to surface. Thus, so-called photo-realistic computer graphic renderers will often employ ray tracing to achieve specular illumination effects, in combination with a diffuse radiation modelling technique based on radiosity computation. While, in principle, it is quite feasible to track specular rays, inferring an illumination model from the diffuse lighting is much more difficult to realise. In fact, it is impossible to infer original light source size, distance, orientation and colour quality with absolute certainty based on diffuse reflections. Nevertheless, this does not imply that we cannot determine some useful approximation.

In the case of diffuse reflection at a point on an opaque flat surface we can, at least, say that there must be an effective light source or sources to the illuminated side of the surface. In effect this narrows the attributable source direction down to a hemisphere centred on the surface point being considered. Considering two infinitesimally small patches,  $\delta A$  and  $\delta B$ , each centred on a separate surface, A and B respectively (and hence having different normals), we could imagine a sphere constructed to enclose the two patches. Any diffuse light reaching  $\delta A$  must have come from the direction of the hemispherical cap enclosing its illuminated side; see figure 1. Similarly, any diffuse light reaching  $\delta B$  must originate from the direction of its enclosing hemisphere. If, for example,  $\delta A$  is in darkness, we can infer that any diffuse light radiation arriving at  $\delta B$  must have come from the direction of  $\delta B$ 's enclosing hemisphere minus  $\delta A$ 's hemisphere, i.e. the wedge-shaped region shaded grey in figure 1. Thus, it appears feasible, given sufficient information from visible surfaces, to narrow down the possible distribution of lighting around the inside of the enclosing sphere in a way that can explain the perceived illumination of the enclosed objects. Any solution that can account for the perceived illumination could, in theory, then be used as a common illumination model that should produce plausible results even though it may not truly match the original light source disposition.

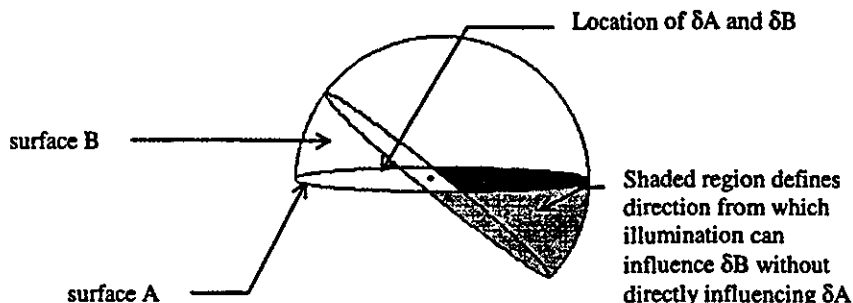


Figure 1

## 2 A Radiosity Solution for Unknown Surfaces

The radiosity method for graphic rendering stems from a formulation first developed to account for heat transfer between surfaces. The approach is based on using the principle of conservation of energy, or energy equilibrium, to find a solution accounting for the radiosity of all surfaces within an enclosure. The original ideas were applied to computer graphic rendering in 1984 when the principles were applied to modelling diffuse light phenomena [8]. Equation (1) describes the situation. Radiosity,  $B$ , is defined as the energy per unit area leaving a surface patch per unit time (i.e.  $\text{Wm}^{-2}$ ) and is the sum of the emitted and the reflected radiosities.

$$B_i dA_i = E_i dA_i + \rho_i \int B_j F_{ji} dA_j \quad (1)$$

$E_i$  is the radiosity emitted from a patch, and the reflected radiosity is given by multiplying the energy arriving at patch  $i$  from all other  $j$  patches in the environment by the reflectivity  $\rho_i$  of patch  $i$ .  $F_{ji}$  is a form factor which expresses the proportion of radiated energy leaving differential area of patch  $j$  and striking differential area of patch  $i$ . It is dependent on respective patch area, orientation and distance.

Using a reciprocity relationship between patches  $i$  and  $j$  and, for a discrete environment, replacing the integral by a summation that assumes constant radiosity over small discrete patches, leads to equation (2).

$$B_i = E_i + \rho_i \sum_{j=1}^n B_j F_{ji} \quad (2)$$

There is such an equation for each surface patch in an enclosure and the complete environment produces a set of  $n$  simultaneous equations which can be written conveniently in matrix form  $Ax=b$ . In normal circumstances, for synthetic graphics rendering, matrix  $A$  is known, as is the vector  $b$ . The  $x$  vector represents the unknown radiosities and can be determined easily using a straightforward iterative technique.

A number of reformulations have been made to the radiosity rendering technique. Most significantly, Cohen, Chen, Wallace and Greenberg [4] developed a progressive refinement approach allowing form factors to be computed 'on-the-fly', hence providing a continually improving image that can be viewed whilst rendering.

In a situation where we have a view of only a portion of the real world, some object surfaces will be visible whilst others, occluded by other entities or facing away from the viewer, will not be seen. In relation to the diffuse lighting within a scene, all these surfaces will have an effect. Other surfaces, outside the field of vision, are also almost certain to be present and, hence, will contribute to the global balance of diffuse radiation. The nature and geometry of surfaces within the field of view may be known or be deduced, but surfaces outside may be unknown and substantially unknowable. This prevents determination of a satisfactory illumination model for the viewed scene. Nevertheless, under certain conditions, it should be possible to infer enough about the global illumination from the intensity across visible surfaces and, hence, to re-illuminate the scene convincingly after the addition of virtual objects or other changes to inter-object geometry. This may be accomplished by enclosing the visible extent of the real world in an imaginary containment sphere which, as a

first step, collects all radiation that cannot be accounted for by the known radiosity of surface patches within the visible scene. After the introduction of virtual objects, or other changes, individual areas on the inside surface of the containment sphere can be treated as emitters to re-illuminate the scene using conventional radiosity principles.

To model this situation, the radiosity equation (2) needs to be revised to account for unknown surface patches. Simply splitting the summation term into three parts does this.

$$B_i = E_i + \rho_i [ \sum_{j=1}^r B_j F_{ji} + \sum_{k=(r+1)}^{(r+s)} B_k F_{ki} + \sum_{m=(r+s+1)}^t E_m F_{mi} ] \quad (3)$$

where;  $r$  represents the number of visible surfaces which are therefore of known radiosity;  $s$  represents the number of invisible surfaces which are known to be within the scene; and

$t$  represents the number of facets within the containing enclosure. (These facets are assumed to have zero reflectivity, but radiosity emittance,  $E_m$ . The sum of the enclosure facet emittances must account for all light radiation that cannot be explained by the surfaces known to be within the scene.)

The first summation represents the contribution to the radiosity of patch  $i$  made by all *visible* patches within the extent of the scene. The second summation represents the contribution to the radiosity of patch  $i$  made by all *invisible* patches within the extent of the scene. The third represents the contribution to the radiosity of patch  $i$  made by all patches making up the containment sphere. Together these must account for all the visible radiosity. Evaluation of the first summation is possible by measurement of the intensity of each visible patch, whereas, the other two summations contain unknown radiosity values that need to be determined.

The enclosed world contains  $(r+s)$  surface patches and, producing an equation for each of these patches, gives  $(r+s)$  equations which will contain  $(s+t)$  unknowns. As for standard radiosity, the problem can be expressed in matrix form  $Ax=b$  and the required solution is given by vector  $x$ .

Unfortunately, things are not so straightforward as they are for standard radiosity computation. The system of linear equations represented by equation (3) are likely to be inconsistent, having no solution that will satisfy them exactly, i.e.  $A$  has no inverse. This is because the form factors relating to facets of the containing sphere will almost certainly not match those of the original light source(s). The sphere's facets will not, except by some extremely remote chance, match the area, orientation and distance of the original sources. In addition, a simple 'best-fit' solution is likely to result in negative radiosities, which is a physical impossibility.

The first problem is overcome by the introduction of an artificial variable into each equation thus ensuring there is a feasible solution; i.e. if all other variables,  $x$ , are set to zero, the vector of artificial variables,  $e$ , becomes equal to vector  $b$ , since  $Ax+e=b$ . The second problem, that of negative values, is addressed by applying range constraints to the permitted solutions. Formulation of an optimised solution results in an objective function that is quadratic, and constraints that are linear, hence, can be represented as a quadratic programming problem.

As any value taken by the artificial variables constitutes a degree of error, the objective function to minimise the error can be expressed as:



$$|e|^2 \quad \text{where } |e|^2 = |Ax - b|^2$$

So, the quadratic programming problem can be expressed as;

*Minimise the objective function:*  $|e|^2$

*Subject to constraints:*  $Ax + e = b; 0 \leq x \leq 1$  where  $Ax = b$  is described by equation (3).

### 3 The Inference Strategy

The overall algorithm proposed here starts with a restricted view of a scene, for which a *a priori* geometric and surface property model has been acquired, and takes the following steps.

- (i) Divide the known surfaces into a suitable number of patches.
- (ii) For each patch, measure its intensity and, hence, determine the radiation incident upon it.
- (iii) Using a standard radiosity model, determine the light radiation incident on each patch that can be accounted for by known patches and subtract this radiosity from that of each patch. (The remaining radiosity must therefore be due to radiated energy from outside the restricted view of the scene.)
- (iv) Completely enclose the scene within a faceted enclosure.
- (v) Determine the enclosure facet radiosities necessary to account for the remaining radiosity within the scene using Quadratic Programming techniques.

### 4 An illustrative example

In this section, the key steps, (iv) to (v), of the above algorithm are demonstrated with an example. The starting point, in this case, is a plane surface object, S, defined by its corner vertices at (3,4,1), (5,4,1), (3,4,-1) and (5,4,-1). The surface is divided into 16 square patches, each of which has an area of 0.25 units. Surface S is illuminated by a plane white light source with boundary vertices at (6,6,-1), (7.414, 4.587, -1), (6,6,1) and (7.414, 4.587, 1). These coordinates put the light source at 45° to the plane of S. The light source is considered to comprise four equal sized facets each of area 1 unit. Applying normal radiosity techniques to calculate the intensity at the centre of each surface patch due to the light source gives the following relative intensity values. To the right of the table in figure 2, these have been converted into actual intensity values, assuming surface S to be white. This forms the starting point of visible, known radiosities and, from this juncture, to demonstrate steps (iv) and (v), we will assume that the original light source is unknown, being outside the field of view.

|        |        |        |        |
|--------|--------|--------|--------|
| 0.0276 | 0.0295 | 0.0295 | 0.0276 |
| 0.0403 | 0.0437 | 0.0437 | 0.0403 |
| 0.0604 | 0.0668 | 0.0668 | 0.0604 |
| 0.0918 | 0.1041 | 0.1041 | 0.0918 |

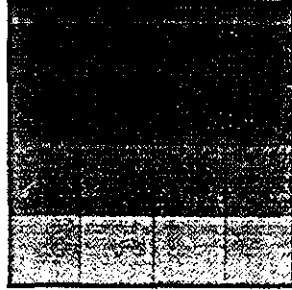


Figure 2

For this example, an enclosure for surface  $S$  is constructed based on subdivision of each face of an octahedron into six triangles, giving a 48-facet enclosure. Each new vertex is then pushed outwards to the full radius of 1 unit and the surface  $S$  is transformed to fit within; see figure 3. The form factors between each enclosure facet and each surface patch are calculated [16]. These are then used in equation (3) with;  $r = 16$  (the number of visible surface patches of known radiosity);  $s = 0$  (the number of invisible surface patches, assuming  $S$  to be single-sided); and  $t = 48$  (the enclosure facets with radiosity to be determined). In this case, we would expect  $A$  to be a  $16 \times 48$  matrix,  $x$  a  $48 \times 1$  vector, and  $b$  a  $16 \times 1$  vector. However, after introducing artificial variables and adding their associated columns to  $A$ , the result becomes a  $16 \times 64$  matrix. Similarly, adding artificial variable rows to  $x$  results in a  $64 \times 1$  vector.

Using a Quadratic Programming algorithm [10] to constrain the 48 enclosure patches to each lie between 0 and 1 whilst minimising the magnitude of the artificial variables, gives the relative enclosure facet radiosities shown in figure 4. (The three enclosure facets determined to account for illumination of the surface in this case are shaded in reverse intensity to their radiosity. Thus, the black shaded triangle corresponds to the brightest facet.) Using this enclosure, and normal radiosity techniques to re-illuminate the original surface  $S$ , gives the result shown in figure 6. Comparison with figure 2 shows that the inferred illumination model reproduces closely the effect of the original lighting on surface  $S$ .

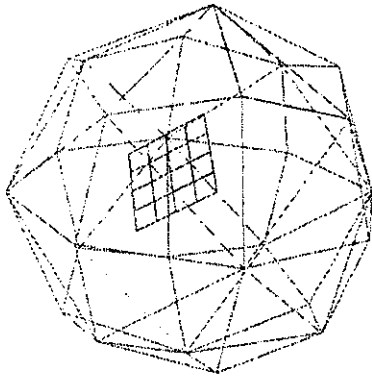


Figure 3

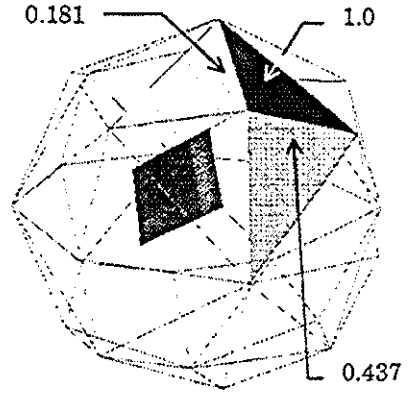


Figure 4

|        |        |        |        |
|--------|--------|--------|--------|
| 0.0290 | 0.0307 | 0.0306 | 0.0286 |
| 0.0416 | 0.0448 | 0.0447 | 0.0411 |
| 0.0612 | 0.0669 | 0.0670 | 0.0606 |
| 0.0915 | 0.1019 | 0.1026 | 0.0908 |

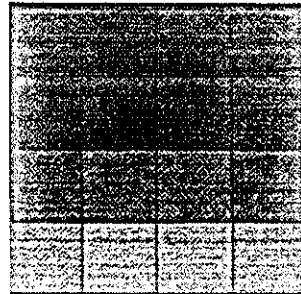


Figure 5

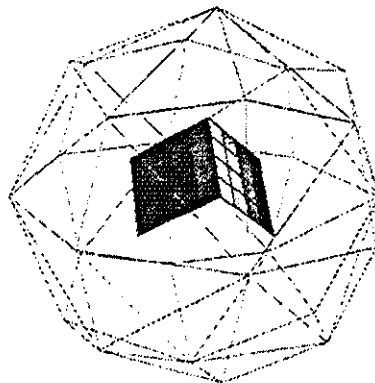


Figure 6

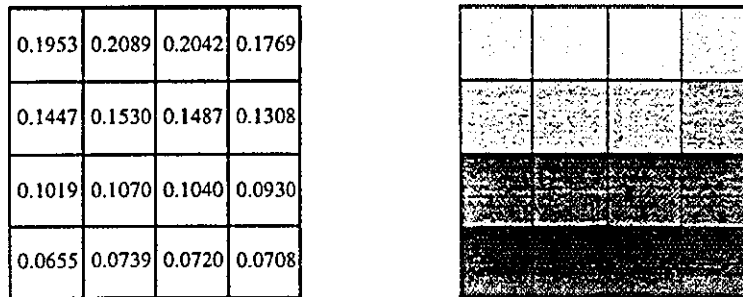


Figure 7

## 5 Conclusion

For the seamless visual integration of entities in mixed reality systems, it will be necessary to develop ways of acquiring common illumination models in environments for which geometric and lighting information is incomplete. This paper represents an attempt to contribute to this end by presenting a strategy for inferring a plausible diffuse illumination model for unbounded environments. The method is based on containment of the known elements within a synthetic enclosure for which a radiosity solution can be found using quadratic programming techniques. Applying the method to re-illuminate a shaded plane surface gives results mathematically close to, and visually almost indistinguishable from, the original. In addition, augmenting the original scene with a new surface produces a visually plausible result. Work is ongoing to determine the optimal relationship between the number of known scene facets and enclosure parameters.

### Acknowledgements

This work was supported by a Coventry University grant. The author would also like to thank; Dr N. Godwin, Professor G. James, Mr N. Steele and Dr J. Tabor of Coventry University, and Professor R. Kalawsky of Loughborough University, for their invaluable help and advice during the course of the work and in the preparation of this paper.

### References

1. Ahlers, K., Kramer, A., Breen, D., Chevalier, P.Y., Crampton, C., Rose, E., Tuceryan, M., Whitaker, R., and Greer, D. Distributed Augmented Reality for Collaborative Design Applications. Proc. Eurographics '95 Conf., Maastricht, Netherlands, August, 3-14, (August, 1995).
2. Bajura, M., Fuchs, H., and Ohbuchi, R. Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient, *Computer Graphics*, 26 (2), 203-210, (1992).
3. Caudell, T., and Mizell, D. Augmented reality: An Application of Heads-up Display Technology to Manual Manufacturing Processes, Proc. Hawaii Int. Conf. System Sciences, 659-669, (January, 1992).

4. Cohen, M., Chen, S., Wallace, J and Greenberg, D. A Progressive Refinement Approach to Fast Radiosity Image Generation. *Computer Graphics*, 22, (4), 75-84, (August, 1988).
5. Feiner, S., MacIntyre, B., and Seligmann, D. Annotating the Real World with Knowledge-Based Graphics on a See-through Head-mounted Display, *Proc. Graphics Interface 1992*, Canadian Info. Proc. Soc, 78-85, (1992).
6. Fournier, A., Gunawan, A., and Romanzin, C. Common Illumination between Real and Computer Generated Scenes. *Proceedings of Graphics Interface '93*. 254-262, (May, 1993).
7. Furness, T. Helmet-Mounted Displays and their Aerospace Applications, *National Aerospace Electronics Conference*, (May, 1969).
8. Goral, C., Torrance, K., Greenberg, D., Battaile, B. Modelling the Interaction of Light Between Diffuse Surfaces. *Computer Graphics*, 18(3), 212-222, (July 1984).
9. Knowlton, K. Computer Displays Optically Superimposed on Input Devices, *The Bell Syst. Tech. Journal*, 56 (3), 367-383, (March, 1977).
10. Lawrence, C., Zhou, J.L. and Tits, A.L. User's Guide for CFSQP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear (Minmax) Optimization Problems, Generating Iterates Satisfying all Inequality Constraints, University of Maryland, Institute for Systems Research TR-94-16r1.
11. Metzger, P.J. Adding Reality to the Virtual, *Proc. IEEE Virtual Reality Annual Int. Symp.*, 7-13, (September, 1993).
12. Milgram, P., Drascic, D., Grodski, J., Restogi, A., Zhai, S., and Zhou, C. Merging Real and Virtual Worlds, *Proc of IMAGINA '95*, 218-230, (February, 1995).
13. Rekimoto, Jun. *TransVision: A Hand-held Augmented Reality System for Collaborative Design*. Sony Computer Science Laboratory Inc., Takanawa Muse Building, 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo 141, Japan, (1996).
14. Rose, E., Breen, D., Ahlers, K., Crampton, C., Tuceryan, M., Whitaker, R., and Greer, D. Annotating Real-world Objects using Augmented Reality, *Proc. of Computer Graphics: Developments in Virtual Environments International 95*, 357-370, (June, 1995).
15. Wallace, J., Elmquiat, K. and Haines, E. A Ray Tracing Algorithm for Progressive Radiosity. *Computer Graphics*, 23, (3), 315-324, (July, 1989).

## Appendix F – Software CD-ROM

CD-ROM containing 3D Viewing and Radiosity software. Written in Delphi Pascal.

The executable file is *rad.exe*. Select **Render** from the *Window* menu to run.

The main scene description must be in *file scene.txt*, with Composite and Primitives in other text files as required.



