Visually Tracked Flashlights as Interaction Devices

Jonathan Green, BSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

April 2008

Abstract

This thesis examines the feasibility, development and deployment of visually tracked flashlights as interaction devices. Flashlights are cheap, robust and fun. Most people from adults to children of an early age are familiar with flashlights and can use them to search for, select and illuminate objects and features of interest. Flashlights are available in many shapes, sizes, weights and mountings. Flashlights are particularly appropriate to situations where visitors explore dark places such as the caves, tunnels, cellars and dungeons that can be found in museums, theme parks and other visitor attractions.

Techniques are developed by which the location and identity of flashlight projections are recovered from the image sequence supplied by a fixed camera monitoring a target surface. The information recovered is used to trigger audiovisual events in response to users' actions.

Early trials with three prototype systems, each built using existing techniques in computer vision, show flashlight interfaces to be feasible both technically and from a usability point of view. Novel methods are developed which allow extraction of descriptions of flashlight projections that are independent of the reflectance of the underlying physical surface. Those descriptions are used to locate and recognise individual flashlights and support a multi-user interface technology.

The methods developed form the basis of Enlighten, a software product marketed by the University of Nottingham spinoff company Visible Interactions Ltd. Enlighten is currently is daily use at four sites across the UK. Two patents have been filed (UK Patent Publication Number GB2411957 and US Patent Application Number 10/540,498). The UK patent has been granted, and the US application is under review.

Acknowledgements

I would like to thank my supervisors Dr. Tony Pridmore and Prof. Steve Benford, for their advice and help throughout my PhD, and for the many opportunities I have had as a result of working with them in the Mixed Reality Lab. In particular I would like to especially thank Tony for his unwavering enthusiasm, patience and support.

Thanks also go to everyone who was involved with, or contributed to, the various different flashlight installations. Most notably: Dr. Boriana Koleva, Dr. Holger Schnädelbach and Dr. Mike Fraser for their work on the various components of the StoryTent and Sandpit; Dr. Ahmed Ghali, Dr. Sahar Bayomi for their work in the Caves; Rachel Fenely for making the 'Journey into Space' event come to life and lastly, Dr. Andrew French, Dr Tony Glover and Dr. Sue Cobb for their work at the Etruria, Magna, Intech and Shepherd School installations. In particular I would like to especially thank Andrew for his help; not only with several of the installations, but also as a very valued and much respected co-worker.

Finally, I would like to thank my friends, family and everyone who ever believed I could do this. Without you, it wouldn't have happened.

Table of Contents

Chapter 1

Introduction

1.1 Vision-based Interaction

Research in the field of computer vision has grown over time a significant sub category of work in the specific use of vision for interaction. Research in this area can typically be divided into two areas of work. The first is the enhancement (or replacement) of traditional techniques for achieving computer interaction (mice, trackballs, joysticks, touch screens etc). These are typically designed for use with the familiar desktop PC. Motivation for new ways of interacting with computers varies from attempts to make them more natural or intuitive to use, to allowing young children (who have been found to struggle with standard interfaces such as mice (Stanton 2001) to learn information technology from an earlier age. A second area where vision for interaction is commonly applied is in pervasive and ubiquitous computing. Here users interact, not directly with computers, but instead with computer-controlled environments or displays. Although present as an essential component of a system, the non invasive, effectively invisible, interface provided by a use of computer vision can be particularly suited to these scenarios. In such circumstances vision interfaces can be either preferable to or utilised in conjunction with other types of physical interface devices such as those using global positioning systems, motion sensors, gyroscopes, wearable computers or magnetism.

Freeman and Weissman replace a TV remote control by having a user wave at a camera while maintaining an open hand gesture (Freeman 1995). The hand can then be tracked in two dimensions to manipulate a cursor over a selection of onscreen controls. Other interactive displays make use of computer vision employing IR cameras and lighting to detect multiple hands or objects placed in close proximity to a semi transparent projection screen (Rekimoto 1997). This

allows a multi touch interface to be constructed and eliminates the requirement for using hand held interaction devices such as IR light emitting pens (Elrod 1992).

Gaming is a rich area in which there is particular interest in developing vision interfaces to replace the traditional console controller. The Sony Eye-Toy is perhaps the best known example (Marks 2004). The Mitsubishi Electric Research Laboratory have also presented a number of ways in which this might be done. These range from recognising hand orientation, together with its relative viewed width, in order to control the direction and speed of a racing car, to allowing users to puppeteer a skateboarder. In this case, the actions of leaning to the left or right control direction (as they would in real life) and additionally ducking and jumping can be detected to allow the skateboarder to avoid high and low obstacles (Freeman, Tanaka Kyuma). Other methods of interaction include recogising arm positions (akin to a child playing “aeroplane”) to control roll in a fighter jet, together with gestures such as lifting both arms together to control altitude. Additionally hands can be used specifically to create a number of recognizable control gestures (Freeman 1998, Kang 2004). These include thumbing left, right, up and down, together with those gestures typically formed when playing a traditional 'rock paper scissor’ game. Hand-based interfaces have also been used to point to select virtual objects (Colombo 2003), and as a three-dimensional mouse (Nesi and Del Bimbo 1996).

The tracking of facial features as two dimensional points in space can also replace the mouse. Gorodnichy et al (2002) detect and track the nose, to create the “Nouse”. Davis and Vaks (2005) present a user interface for a responsive dialog-box agent that uses real-time computer vision to recognise user acknowledgements from head gestures, where a nod means “yes” and a shake means “no”. El Kaliouby and Robinson (2003) describe a similar affective message box, which employs a real time gesture recognition system as its input modality.

Kawato and Ohya (2001) propose an approach for detecting nods and shakes in real time from a single colour video stream, which depends on detecting and tracking a point between the eyes. Kapoor and Picard (2001) describe a system that detects head nods and head shakes in real-time using an infrared sensitive camera equipped with two concentric rings of infrared LEDs to track participants' pupils, and eye tracking is also employed in the system proposed by Tang and Rong (2003). Morimoto et al (1998) employ an explicit three-dimensional model, describing the participant's face as a planar surface and basing the recognition of head gestures on changes in the parameters of that plane. The plane representation is only a very crude approximation to the human face and captures only a small part of the facial variation that takes place during conversation. Moreover, the face must be (almost) entirely visible if a plane is to be fit with the necessary degree of accuracy.

Full tracking of body part movement can allow for the control of a 3D spatial sound system (Bradski 2002). Analysis of movement between poses is used to trigger response such as positioning, starting, stopping and controlling the tempo of sounds. Another common use of vision for interaction, specifically in 3D environments, is in recovering the position and orientation of fiducial markers (Kato 2000). By holding and moving such markers, these effectively become devices by which users can manipulate virtual objects in augmented reality applications. Fiducials can be attached to almost any object, creating a wide variety of vision-based tangible interfaces.

1.2 Interacting with Flashlights

Flashlights are particularly interesting devices upon which to build human-computer interaction technologies. They are cheap, robust, fun as well as being available to, and understood by, a sizeable user base. Most people from adults to children of an early age are familiar with flashlights and can use them to search for, select and illuminate objects and features of interest. Flashlights for example

have been used specifically as an intuitive means to illuminate virtual objects in an augmented reality environment (Regenbrecht 2002). Here, it is not the real beam of a flashlight that is utilised for interaction purposes but instead the flashlight itself becomes a recognisable prop to metaphorically represent the location of a virtual light source. Using fiducial markers attached to the flashlight (which are monitored by a camera attached to a user's head mounted AR display) it is possible to have a virtual light source follow the flashlights movements in 3D. To the user it appears as if the real flashlight is shining light into the virtual world.

One way to exploit the actual beam of a flashlight is to attach light sensors to a surface and use flashlights in order to activate them. An interesting variant on this makes use of a modulated beam, which is capable of carrying identification signals, so that only certain sensors are activated when illuminated. This has been used as a method for tagging and finding individual books in archives (Hongshen and Paradiso 2002) by sweeping them with a 'flashlight like' defocused laser beam. One potential modification might be to allow tags to respond differently to different signals (e.g. associated with different 'flashlights') thus allowing the association of different content. Tagging or use of light sensors on surfaces can however be infeasible if the surface is fragile, valuable or inaccessible. This can greatly reduce the range of locations in which such an interface might be deployed. Difficulty of reconfiguration, battery life or provision of power/cables to sensors can also present issues.

1.3. Visual Detection and Recognition of Flashlight Projections

Flashlights have the potential to provide surprisingly rich vision-based interfaces. The area(s) illuminated and the motion of the flashlight beam across the physical world provide valuable information regarding the user's interests and intentions. The projection of a flashlight beam onto a physical surface varies considerably with the physical structure, position, and orientation of the flashlight and is

generally visible in images captured using standard equipment. This raises the possibility of both recognising individual devices and recovering their properties from images of flashlight projections. It is the aim of the work reported here to investigate this possibility.

Interfaces based upon visual detection and recognition of flashlight projections have a wide variety of potential applications. Flashlights are available in many shapes, sizes, weights and mountings. Tightly focused, hand-held pencil flashlights can be used in the detailed examination of small features and objects. In contrast, floor mounted searchlights can be used to illuminate large sections of, e.g., buildings. Flashlights are particularly appropriate to situations where visitors explore dark places such as the caves, tunnels, cellars and dungeons that can be found in museums, theme parks and other visitor attractions.

Many sites in the museums and heritage sector employ mobile audio guides that rely on handheld computers to deliver context sensitive audio information to museum visitors (Aoki and Woodruff). Some of these utilise position tracking to deliver location dependent audio and others allow proximity based interaction, for example via RFID. Related products have some of the attributes associated with flashlights. Localised pointing can be achieved, for example, by aiming infrared beams at special beacons. As with light sensors (or light sensing tags) however, the latter typically involve attaching objects to a surface and also tend to have a limited reading range and/or lack of high spatial resolution. This makes it difficult to accurately point at targets over a significant distance. Visual detection and recognition of flashlight projections has the potential to overcome these difficulties, and play a valuable role in museum experiences.

Flashlights are also suited to interacting with technology outdoors at night. Stronger flashlights can be used in more brightly lit situations, e.g. when interacting with projected graphical displays. In larger spaces it is natural for several flashlights to be used simultaneously. This provides interesting opportunities for group interaction.

Perhaps most closely related to the concept of an interactive device based upon detection and tracking of flashlight projections is the use of laser pointers to remotely manipulate graphical objects on large shared displays (Davis 2002, Olsen 2001). Laser pointers potentially allow fine pointing and manipulation of objects and seem a highly appropriate technology for meeting rooms, lecture halls and similar environments. Given that visually tracking laser pointers is such a closely related technique, it is worth examining the differences between flashlights and laser pointers and, in particular, the interesting characteristics that might make flashlights especially suited to use in certain public settings.

- Instead of a point of light, the recoverable image of a flashlight beam on a surface takes the form of a pool of light meaning it is possible to indicate an entire region at any point in time. This means a user could use it to sweep out or interact in varying manners with targets of different sizes.

- This pool of light cast on surfaces by flashlight beams contains significant information that might be exploited by a tracking system. Of note are its shape, which varies according to the flashlight's orientation in relation to the target surface and its size which varies with distance. Additionally, different flashlight beams can exhibit different patterns of light intensity. This makes them potentially identifiable.

- Working with pools of light opens up opportunities for collaboration. For example, it may be possible for a system to respond differently depending on how much of a surface or target is being illuminated i.e. use of several beams in close proximity to cover a larger area

- Flashlights can be safely given to children for unsupervised use. There is no danger if they are shone into eyes and additionally they are unlikely to cause damage to surfaces.

1.4 Thesis Structure

This thesis examines the feasibility, development and deployment of visually tracked flashlights as interaction devices. The location and identity of flashlight projections is recovered from the image sequence supplied by a fixed camera monitoring a target surface. The information recovered is used to trigger audiovisual events in response to users' actions. Having considered similar and related work with regard to vision based interfaces, the remainder of the thesis is organised as follows:

Chapter 2 reports early trials with three prototype flashlight interfaces, each built using existing techniques in computer vision. Analysis of these trials shows flashlight interfaces to be feasible (both technically and from a usability point of view) and that participants found them attractive and enjoyable to use.

Though the techniques employed in early systems allowed the creation of interesting and usable interfaces, they were not designed for use with images of flashlight projections and so suffer some limitations. Chapter 3 therefore addresses the problem of extracting stable description of the patterns of light projected by flashlights from images of those projections. Several methods are developed and evaluated, and a novel quotient-based method selected for subsequent use.

Chapter 4 discusses the deployment of flashlight interfaces in a number of different situations, highlighting each location's technical issues and challenges. The environments considered include a laboratory-based demonstration rig, an effectively outdoor installation, an interactive storytelling environment involving groups of users, a multi-station interface using standard flashlights, and a large scale installation employing searchlights.

The techniques used to provide an interactive technology suitable for use in these diverse situations are presented and examined in detail in Chapter 5. Interactive flashlight systems have been commissioned in each of the scenarios considered.

Some of the issues raised in Chapters 4 and 5, however, cannot be resolved. These are presented, with consideration of their effects, in Chapter 6. Finally, future work is discussed in Chapter 7.

The methods developed here form the basis of Enlighten, a software product marketed by the University of Nottingham spinoff company Visible Interactions Ltd. Enlighten is currently in daily use at four sites across the UK. Two patents have been filed (UK Patent Publication Number GB2411957 and US Patent Application Number 10/540,498), and a patent search was undertaken by the patent office. The UK patent has been granted, and the US application is under review.

1.5 Contributions

The work presented provides scientific contributions to both the fields of Human Computer Interaction (HCI) and Computer Vision. To HCI, a new method of interaction is presented, with initial trials indicating a strong suitability for its use with display based educational material, or locations/exhibits in museums and heritage centres. Insight is provided into how such an interface is deployed as well as observations regarding its use in different contexts and by different age groups. To Computer Vision, a novel image processing technique, the Quotient operator, is presented which is suitable for the adequate extraction of projected flashlight profiles from captured images, for use in recognition. A full discussion of the implementation of a flashlight based user interface, that utilises such an operator, is provided.

1.6 Summary

In this chapter we have briefly reviewed vision-based interactive devices. We have considered the advantages of flashlights as devices upon which to build such an interface. We have presented the aims of the work reported here, which are to develop, deploy and evaluate vision-based interactive devices centred on standard domestic flashlights. The structure of the remainder of the thesis is described.

Chapter 2

Early Installations, User Reactions and Issues

Analysis of the relevant literature suggests that standard domestic flashlights might form the basis of a natural interactive device useful in applications such as interactive exhibits in museums and activity centres. To assess the feasibility of visually detected flashlight projections to support such a device, a number of initial systems were constructed using standard methods and techniques from computer vision. This chapter focuses on three such installations and the computer vision methods involved. Users' reactions to the prototype flashlight interfaces are discussed in each case, and technical and usability issues are identified.

2.1 The StoryTent

2.1.1 Introduction

The first experiments with interactive flashlight technology were carried out as part of a larger body of work researching the development of interesting and novel collaborative storytelling environments. Such environments have been shown to have both educational and social development benefits for young children (Bayon 2003). Previous research reported at CHI has discussed using a tent as a projection interface for ambient and informal experiences (Waterworth 2001) which might be suitable for this purpose. In exploring the use of tents, work was carried out to develop such an interface with the aim of giving young children an engaging and shared experience of a virtual world. Such an interface could be deployed in public spaces such as museums, theme parks and classrooms (Green 2002).

The StoryTent (as the interface became known), in its most basic form, consisted of a fabric projection screen stretched over a lightweight aluminum tubing frame

to mimic the shape of a classic A-frame tent (Fig. 2.1). On either side, two projectors, outside the tent, displayed views of a common 3D virtual world onto each surface. These were positioned and synchronized (using the MASSIVE distributed VR system: Greenhalgh 1995) so that those inside the tent appeared to be looking out into a 3D virtual world.

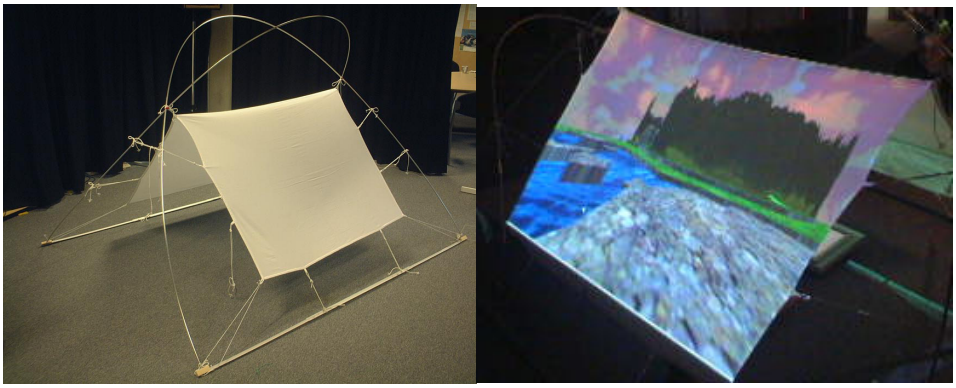


Fig 2.1 – The A-frame StoryTent

To control the environment, focus was placed on interactions that fit naturally with the tent metaphor. In keeping with this theme, the concept of flashlights aimed at the surface of the tent, and visually tracked from outside, was proposed as a suitable interaction technique to aid with the story telling experience. Use of flashlights in this way has a number of advantages. Like the tent, it fits the theme of a camping experience. From a usability point of view, flashlights are light weight, familiar and easy for children to control. Use of multiple flashlights simultaneously also opens up possibilities for group collaboration. Finally, the interface is advantaged in that it can be used by participants who are both inside and outside of the tent, by aiming flashlights at either surface. Interactions can be easily viewed from either side.

In addition to the use of flashlights for interaction, the tent also incorporated another technology which consisted of a modified and enlarged RFID tag reader, built into one of the tent's two entrances. As shown in Fig. 2.2i, this reader formed an obvious doorway through which all entrance and exit must occur. The doorway emitted high frequency radio waves (13.56MHz) which induced a response in tagged objects that passed through it allowing them to be recognised.

Prior to trials, several virtual objects associated with specific tag IDs were prepared and these tags were attached to real life versions of the virtual objects. This allowed designated items, brought into the tent, to not only have a physical presence within the tent itself but, additionally, the virtual equivalent of the physical object would also appear within the virtual space that the tent inhabits. Due to signal strength issues and the unpredictable manner in which such objects could pass through the entrance to the tent (carried in hands, pockets, bags or even thrown), each object needed to have multiple tags attached in orthogonal orientations as shown in figure 2.2ii.

i) ii) iii)

Fig 2.2 – i) The RFID tag reader doorway ii) Orthogonally tagged physical balloons iii) Associated virtual balloons

For trials with flashlights, a number of different coloured balloons were prepared and these, when taken into the tent, would cause their equivalently coloured virtual balloon to appear to fly over from the distance and hover near a surface of the tent (Fig. 2.2iii). Participants could then use flashlights, directed at the surface of the tent, to control an individual balloon's movement in the virtual sky. The balloons would follow the flashlights in exact correspondence with the surface of the tent and this created an effect as if the two were connected by invisible elastic (Fig. 2.3). Turning a flashlight off would of course sever this connection, returning the balloon to its default position. Since the flashlight part of the environment needed to therefore have knowledge of the items present within the tent, and additionally control the graphics associated with their virtual versions, a method of integrating the three elements was required. To achieve this, a multi-

platform communications architecture called EQUIP (Greenhalgh 2001) was utilised, as this allows different components of a distributed system to communicate via a centralized database.

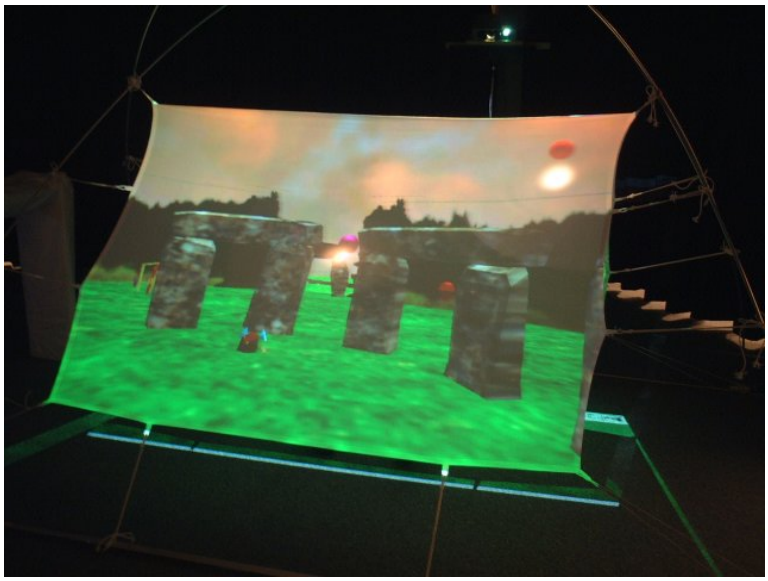


Fig 2.3 – Balloons following flashlights

In addition to its use as a trial application for initial development of and experimentation with flashlight technology, the StoryTent also reflected several concerns within HCI. Firstly, as a result of being a closed space where entry and exit occurs by means of a single portal, it represented an example of a traversable interface that provides the illusion of crossing into and out of a virtual world. Previous examples of such interfaces have included fabric curtains, sliding doors, hinged screens and even water sprays (Koleva 2000) however the StoryTent demonstrated an alternative. Here, while participants still entered an area that is defined by the projection screens, unlike the obviously comparable CAVE-style immersive interfaces, the space outside the screen remained part of the experience. The tent also supported effective collaboration which is a concern for the designers of children's storytelling technologies. Previous solutions include the use of single display groupware with multiple input devices (Stewart 1999), or room size projection systems combined with physical and tangible interfaces (Bobick 1999) however the tent met this requirement by way of the two interaction styles described. Additionally, since adding new flashlights and

balloons required no changes to either the hardware or software, the system was inexpensive and easy to configure. As a final point, the tent was designed to meet some of the challenges of designing interfaces for public spaces for example, the studies of interactive exhibits in museums which show how passers-by can learn by watching others interact (vom Lehn 2002). Since the two-sided nature of the tent granted those outside a public rendition of the activity going on inside, while at the same time maintaining a relatively protected and isolated environment for those inhabiting the tent (Fig. 2.4), this requirement was also successfully met.

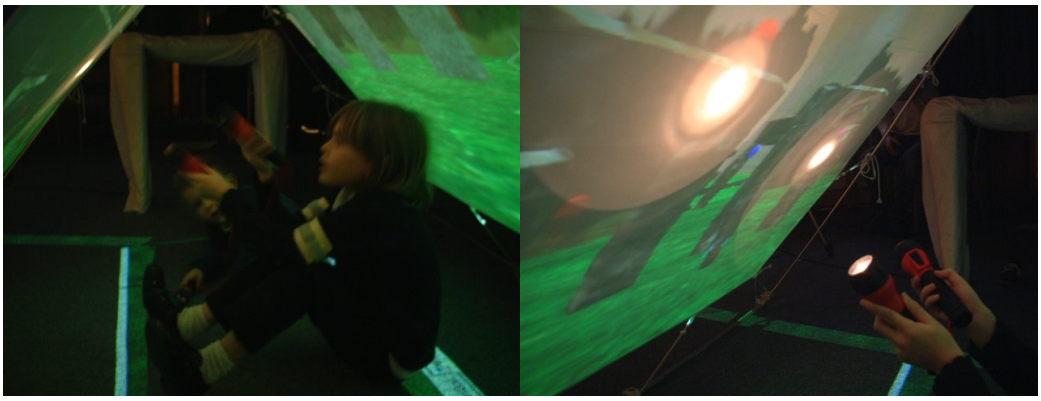


Fig 2.4 – The environment inside the tent and using flashlights on the inside surface

2.1.2 The Flashlight Component

As an initial specification, the flashlight interface, to be used with the tent, was to be able to detect and track multiple beams on each surface which might be controlled by users who are either on the inside or outside of the tent. Flashlights were to be individually followed as they moved from one side of the tent to another (via the ridge) and additionally, the system needed to be robust enough to cope with fast movement, momentary occlusion and beams potentially interacting (overlapping) with one another. The physical setup used to achieve this comprised cameras mounted on top of the tent's projectors. This placed them not only high up out of the way, but also ensured that the principal axes of the camera and projector were approximately parallel. Both cameras were zoomed in to view exclusively the pattern projected into the tent. The projections used were to be non-static (due to the moving virtual objects) and unconstrained in content. For

this reason, during testing, a view out into a 'Stonehenge' world was to be used as this was deemed to contain a good range of likely colours and textures of the kind that might be present in any future applications. An ability to follow flashlights against such projections, without being distracted by their content, was also a requirement of the interface.

2.1.3 Developing a Flashlight Interface

Development of a flashlight interface, capable of achieving the above outlined goals, required answers to three questions. First, how can we detect and separate light resulting from a flashlight beam from patterns produced by a projector? Second, what was a good way of following targets, that have very unconstrained and unpredictable movements, in real time? Finally, how can the output of such an interface be integrated with and control the virtual environment that it is being developed for?

Considering the former of these questions the problem at first appears trivial with the immediately apparent solution being to simply look for the brightest point in any given scene and apply a threshold. Such a solution is likely to be successful. However, the tent situation was found to be problematic for the following reasons:

1. It was not a direct light source that was to be detected but instead, merely its effect on a partially transparent surface
2. The flashlight illumination was not measured on its own but was instead mixed with that forming a projected virtual background
3. Flashlights may not have been (and often were not) the brightest regions in the scene
4. Projections ideally require dark or dimly lit rooms in order to be seen clearly. This is less than ideal for recovering a good dynamic intensity range from camera inputs

In an attempt to address some of these environmental factors, the detection method used in this first flashlight system combined both thresholding and shape analysis. In brief, for each frame input to the system, images were converted to represent intensity and these values were thresholded at a level chosen in order to minimize noise. In the presence of flashlight beams, or particularly bright areas of the projection, the resultant binary image would commonly feature a number of regions to which contours were fitted. Each of these would then be analysed in turn and, based on the known constraints of the system (in this case small torches held close to a surface), a size threshold was set. This allowed unusually sized bright regions to be disregarded (which were unlikely to contain flashlights), and the remaining candidates were passed to a secondary stage of processing.

Each flashlight that might be used is associated with a tagged balloon, and the tag reader informs the system when each balloon enters the tent. The number of virtual balloons to be manipulated by flashlight projections is therefore known. This places an upper limit on the number of flashlight projections to be detected and tracked by the system. In this early installation a simple detection algorithm was employed. When seeking n projections, the system merely scans the image from top-left to bottom right and assigns the first n appropriately sized, supra-threshold regions to the n balloons present in the tent. Though this can cause some miss-assignment when flashlight-like bright regions of the projected image appear in the upper left, it was not a significant problem.

Once a balloon tag was associated with an image region, that region was tracked to maintain its identity between frames. A two-frame data association algorithm was employed. A minimum enclosing circle was fitted to each region and the circle centre taken as a crude approximation to the centre of that region. The centres of the regions associated with a balloon in the image captured at time t-1 were projected, using a simple constant velocity motion model, into the image captured at time t. A circular area centred on the projected location and of radius equal to the speed of the projected region was examined; any newly identified

regions whose centres lie in this area were considered to be potential matches (Fig. 2.5)

If only one candidate was found to be located within a particular predicted region, the flashlight associated with that region was updated with that candidate's position. If more than one candidate was found, the one closest to the flashlight's predicted centre point provided the new position. Should no candidates be found, a previously recognised balloon clearly could have its position updated. To deal with this, a system of expiry time was used so that flashlights, failing to be updated in a single frame, were not considered lost until they were not updated for a period of *n* frames. Any newly found regions (at time *t*) that were not associated with a balloon region (from time *t*-1) were considered to be new flashlights that had only just been turned on.

i) ii)

Fig 2.5 – Predicting the movement of flashlights. In i), Flashlight 1 is moving slowly and 0 is stationary. In ii) both are moving. The line from the center of each marked beam indicates its most likely location next frame based on known speed and direction. The circles around these locations indicate the margin of error. Flashlight 1 is moving faster hence will accept matches from a larger area.

Flashlights can move quickly and unpredictably, in any direction at any time, making it hard to identify an optimal tracking solution. An alternative might have been to utilise a general tracking engine such as Kalman filtering (Kalman 1960), particle filters (Isard 1998) or kernel mean shift (Comaniciu 2003). Other, application-specific methods e.g. those used to analyse the (comparably fast) movement of snooker balls (Denman 2003) may also have been applicable here. Many application specific vision solutions (e.g. Gao 2001, Heap 1995, Rehg

1994), however, exploit the constraints of their situation to an advantage ("*the application often constrains the vision problem to be solved",* Freeman, 1999). When following someone over a view of a car park for example (Khan 2006), the tracker could take advantage of the fact that the subject will not suddenly sprint to the middle of the frame. In the case of the StoryTent however, no strong restrictions could be applied. Although the discussed (and other) tracking methods could have been experimented with, the chief aim of the early work reported here was to assess the value of the interactive flashlight concept. Since the algorithm used was sufficient to allow such assessment to be carried out, employment of more complex techniques was not necessary.

Integration of the flashlight interface with the other technologies present in the tent primarily involved two issues; namely communication and calibration. As noted, the elements contributing to the StoryTent environment were designed to be integrated using the EQUIP platform. This was found to work well except in situations where many flashlights were being used. In these situations, during development, the fact that each flashlight was being updated at 30Hz would often cause some updates to be lost during transmission between the flashlight interface and EQUIP. This typically led to sporadic and random positioning of the balloons within the virtual display. The solution was to apply an element of filtering to reduce the number of updates required to be sent. This required balloon positions to be interpolated so that they followed smoothly behind the flashlights controlling them. While, for most, this was found to create a pleasing effect, with some younger users (see section 2.1.4), it proved to make the system too unresponsive.

Calibration, in order to achieve correspondence between flashlight positions and the projected positions of balloons within the virtual world, also proved to present a significant challenge. In order to provide the illusion that a balloon is following a flashlight, the two must move exactly or as close to each other as possible and this must be the case over all, or as much as is possible, of each tent surface. The solution, which also allowed for following of flashlight beams as they moved over

the ridge of the tent, involved changes to both system. To provide relative tent coordinates, each frame was warped so that the corners of the tent corresponded exactly with those of the image corners and were combined to align along the apex of the tent. This meant that following a flashlight, as it moved from one projection surface to another, was no different from following it on a single surface. The coordinates of found flashlights, which were now aligned exactly to the tent surface, were then passed through a translation matrix and this converted them to equivalent balloon positions for use in the 3D world.

2.1.4 User Trials and Issues

Despite the techniques and methods used to create the StoryTent's flashlight interface being standard and simplistic, many of the problems discovered during trials were related more to the design of the test content than the actual interface. The trials carried out involved having two children of ages 7 and 4 interact with the tent using flashlights from both inside and outside. The children were given no instruction as to what was expected of them but soon picked up the link between flashlights (found in the tent) and the movement of the balloons on one side (see video). They were then asked to perform simple tasks such as moving the balloons round the edge of the tent surface and using flashlights both outside and in the tent simultaneously.

On the whole results were promising but a number of issues arose. As regards control, the younger of the two had problems with slow navigation of balloons, instead preferring to wave the torch erratically. This caused problems in his understanding of the interface due to lag. As mentioned before, the balloons acted as if they were connected to flashlight beams on a band of elastic which, from an interaction view, mostly worked well. A downside of this is that, when the flashlights were moved fast, the balloons were found to follow too far behind, making them feel as if they were acting entirely independently. This broke down any chance of an association between the two for a child. Together with the slow

speed of balloons, another contribution to this lack of understanding occurred when a flashlight failed to be correctly detected and tracked. This was largely due to issues of detection in the system. These were caused either by poor configuration of system and camera parameters or occlusion of the cameras, making detection impossible.

Configuration of cameras was difficult because they were focused on variable, often bright surfaces. Little image contrast was present between the projected display and the flashlights used. In order for flashlights to be apparent against such a bright surface, it was necessary to adjust the cameras so that images appeared very dark. Since this was commonly found to cause flicker in the resultant output, however, a compromise was required between minimising this effect while not saturating the brightest regions of the projection.

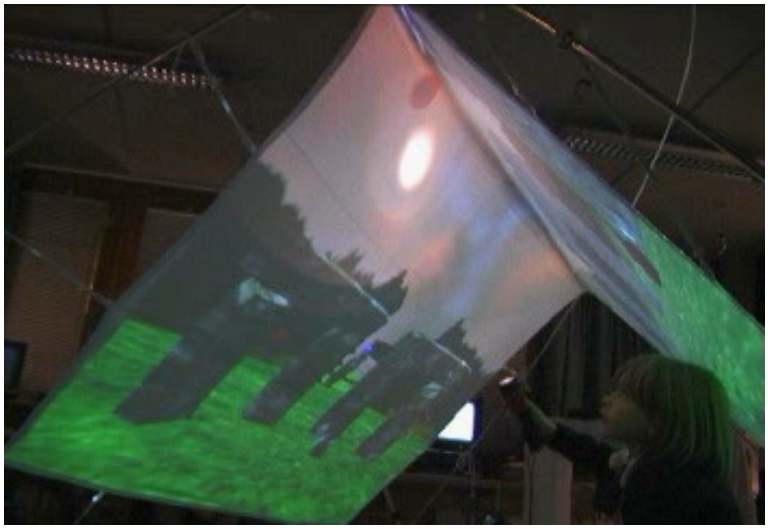


Fig 2.6 – Testing fine control with flashlights by moving a balloon around the edge of the tent surface

The final problems encountered were largely down to the design of the trial. For example, the children often found that the balloons they were controlling were swapped with each other's. This caused confusion as, once a balloon had been acquired, the natural impulse observed was to focus on that balloon and that balloon only. As safeguards were built into the system to prevent overlapping beams being confused swapping, again, was caused by a loss of detection. Since balloons were reassigned based on a predefined order, once tracking of the lost

flashlight was resumed, it may not have been assigned the same balloon that it had originally been controlling.

A related problem was found in the design choice that balloons which were no longer following flashlights would always return to the centre of the tent surface. This caused much confusion when the eldest of the children was attempting the task of moving a balloon about the edge of the tent (Figure 2.6). Since it was not immediately apparent that moving the flashlight too close to the edge meant a loss of detection, this would commonly cause the child to have to reattempt the exercise from the start. A solution to this problem would have been, either to only allow balloons to be controlled once a flashlight gets near them or, alternatively, find a way of always assigning control of the same balloon to the same flashlight, potentially via some form of recognition.

In summary the main issues found with the StoryTent flashlight interface were:

- Loss of flashlight detection due to the mechanism used and the difficulties of camera configuration in configurations that contain projections

- Speed of response to actions in some cases being too slow for a user to be able to understand how the system is working

- A need for a stronger association between flashlights and the things that they control

To conclude, the trials with the flashlight element of the tent environment suggested it to be a workable technology. This led to its conversion for use in further trials as part of a different experience, as discussed in the next section.

2.2 The Sandpit

2.2.1 The Living Exhibition

The original trials with flashlight technology in the StoryTent environment, despite some issues, were found to show great potential. In order to further test the technology, work was carried out to adapt it for use in larger scale public trials. To this end, the flashlight interface was featured as part of a large event known as 'The Living Exhibition' (Ng, 2002).

The Living Exhibition was a four day event held in the grounds of The City of Nottingham's Legendary Castle and run by the University of Nottingham's Mixed Reality Laboratory in collaboration with the tourist attraction's curators. The event was designed with mass participation in mind and was intended primarily to appeal to families. The aim of the exhibition was to make a visit to the Castle a more fun and informative learning experience and with this in mind, was built around the concept of a classical treasure hunt.

Upon registration for the event, tourists and guests were given information packs and encouraged to tour the castle and its surrounding area where upon they looked for clues or hints as to the history of the fortress. Typically this involved visiting a number of set locations, answering questions and doing sketches or brass rubbings of things that could be found there. These would be brought back, scanned and RFID tagged (see section 2.1.1) to associate them with their relevant locations, allowing them to be used with a collection of technologies, so that further information about each location could be found.

The technologies available at the Living Exhibition ranged from the Augurscope (Schnädelbach 2002) (a portable device that used 3D computer graphics to allow users to see the Castle layout as it was in the past) to a customised implementation of the StoryTent, the original form of which was discussed in the previous section. At this event however, interactive flashlights (in their original form) were not

used as part of the tent experience. Instead, these were employed as a control system for another technology - the Sandpit.

Fig 2.6 – i) Using flashlights to displace sand in order to uncover images buried in a virtual sandpit ii) Background: Users selecting drawings to place on a tray of real sand to create 'diggers'. Foreground: Children aiming flashlights into the virtual sandpit. Rocks and sand around projection edges were used to enhance the illusion

The Sandpit was an interactive exhibit designed with the aim of allowing users not only to experience their own content in a new and interesting way but also to discover new things relating to it. The exhibit, as its name suggests, took the form of a graphical representation of a sandpit, projected onto the floor, with which users could interact by digging in the virtual sand as they might in a real sandpit. Upon arrival, users would place their drawings on a separate tray filled with sand (Fig. 2.6ii) whereupon two (graphical) sparking 'diggers' would appear and start to move randomly across the surface of the pit (Fig. 2.7). These acted as a trigger in order to direct the participant's attention towards the Sandpit. Beside the Sandpit were placed two flashlights and these, when aimed at its surface, allowed the users to control the diggers to displace the virtual sand. Users would work with the flashlight by sweeping the surface of the pit looking for images that were buried in the virtual sand. When uncovering something, directing the flashlight persistently over that region and its surroundings would slowly displace the sand till the image was exposed (Fig. 2.6i). Once done this would appear to spin upwards while growing in size so that it could be clearly seen and then vanish after a short time (Fig. 2.7). These images would include the participants own

drawings (scanned and tagged earlier) and other photographs that were related to their content.

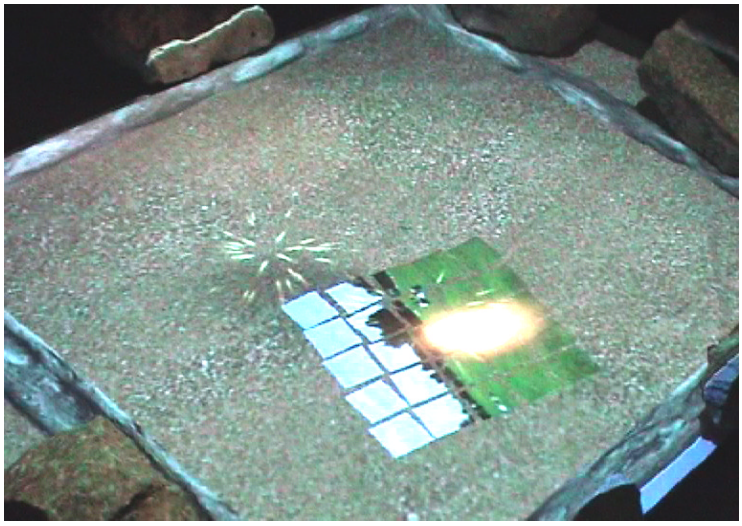


Fig 2.7 – Sparks moved randomly over the surface of the pit to indicate available 'diggers' that could be controlled via flashlights. Once found, images would begin to disintegrate and vanish after a short time,

2.2.2 The Flashlight Component

From a development point of view, the Sandpit comprised much of the same technology used within the StoryTent. For physical construction, a projector was attached to an overhead gantry and this was deliberately aimed, using as steep an angle as possible, so as to create a projection on a screen on the floor that would not be easily occluded by users or flashlights. To disguise the projection screen, rocks gravel and sand were used around its edge (Fig 2.6ii). This created the effect that a hole in the floor had been created and hence added to the authenticity of the experience.

In this scenario, tagged objects did not provide targets for manipulation with the flashlights but instead were used to allow a participant's drawings to be identified. The reader (this time hidden within the aforementioned tray of sand) identified the tagged paper that was placed above it and this was used to retrieve the digital scan of the same drawing, together with any other photographs related to its content.

These were then hidden within the virtual sand ready to be discovered using flashlights.

Programmatically, flashlights in the Sandpit were employed in much the same way as they were within the StoryTent i.e. to control an entity within the virtual world over a two dimensional space. As noted in section 2.1.3, use of EQUIP for communication between system elements had proved problematic for high frequency positioning updates. For the Sandpit therefore, standard sockets were utilised and this allowed the diggers to move in direct correspondence with the positions of the flashlight, at a speed synonymous with that caused by the user. The computer vision methods employed in the Sandpit version of the flashlights system were identical to those used in StoryTent.

2.2.3 User Trials, Observations and Issues

During the Sandpit trials, although the flashlight interface used was similar to that employed within the StoryTent, its application and style of use was observed to be very different, even with young children. In situations where their flashlight failed to be associated with a digger, for example, some children failed to realise that anything had changed, and continued to use the flashlight as usual. Unlike the StoryTent, where most breakdowns were a result of failures in flashlight detection, the Sandpit more commonly suffered from issues relating to the system's tracking element. This was due to the unpredictable nature of the content found within the sandpit. As participants' scans (primarily white in colour) were uncovered, these effectively created bright regions in the projection and these, with the detection method employed, appeared very similar to the flashlights being used to uncover them. Sometimes, if the tracker began to 'follow' such images instead of flashlights, these were observed to, in effect, self excavate. This problem was similarly observed to occur when the bright bottom of the Sandpit itself was uncovered or occasionally (due to poor calibration) when the sparking diggers themselves were mistaken for flashlights. In each of these cases the problem was

commonly rectified by effectively 'collecting' the tracker by moving a real flashlight over the area that was being mistakenly tracked.

The remainder of the observations during the Living Exhibition trials with the Sandpit were primarily related to issues of design and access. As noted, access to the pit was deliberately limited to a single side as it was thought that this would be a minor drawback compared to the gains it provided in minimising occlusion (see section 2.2.2). Such restricted access in fact turned out to have a larger effect than expected as commonly it was observed that users would only explore the near side of the pit, causing them to find few or no pictures. This was particularly apparent in the very young (due to height) and more elderly users (due to a desire not disrupt the projection and hence their experience of the exhibit). In fact the disruption caused by leaning too far over the projected area acted as a strong deterrent, meaning that occlusion was rarely an issue. It is therefore likely that a straighter projection, that permits access from all sides, would be of benefit.

The remaining design issues observed were largely simple oversights, such as images appearing upside down or disappearing after too short a time. This made it hard for them to be recognised by users. In conjunction with this, however, some users commented that they would have liked to have been able to control images once unearthed (similar to the click and drag action of a mouse). This presents a question as to whether or not flashlights could also be used as rotation devices as such an action would also allow the correct orientation of upside down images (see Chapter 7). In effect, rotating the flashlight would rotate the image under its control. Similarly, potential was observed for the exploitation of natural gestures. For example, wiggling a beam in circles, in an effort to dig harder, was observed frequently throughout the trials. Additionally, users (in particular children) would often realise when another had found something and would move their flashlight to the same location in an effort to 'aid' the digging or 'dig harder'. This opens up possibilities for collaboration.

In summary the main issues found with the Sandpit flashlight interface were:

- Disruptions to tracking due to confusion between flashlights and projected objects of similar size and intensity

- Reduced or misunderstood exhibit usage due to limited access areas and a desire not to disrupt the display

- Expectations of or suggestions for unimplemented functionality such as gesture recognition, collaborative use of flashlights or an ability to control images and their orientation.

To conclude, the extensive public trials, carried out over the four days for which the exhibition was run, again showed interactive flashlights to be a viable technology. During the same exhibition, trials were also carried out utilizing a different instantiation of the technology. Details of changes to the flashlight interface, and the new style in which it was used, are described in full in the next section.

2.3 The Nottingham Caves

2.3.1 Introduction

During the latter half of The Living Exhibition (see section 2.2.1) a second trial, featuring a different use and instantiation of the basic interactive flashlight concept, was undertaken. This transposed the technology from its use with projected surfaces to natural ones, this time found within an underground cave environment.

Fig 2.8 – Guided tours into the Caves under Nottingham Castle

Beneath the grounds of Nottingham Castle are a number of man-made caves that are rich in historical information and often feature in a visit by way of guided tours where exciting stories, conveying the history of the caves, are recounted by a trained guide (Fig. 2.8). Over the seven hundred years since they were first carved out, the caves have served several different purposes, ranging from use as store rooms, larders and prisons (allegedly holding King David II of Scotland for eleven years) to secret passageways which, on one occasion, were used for the kidnapping of the Queen of England. During the exhibition one such cave, known as 'King David's Dungeon', was used and converted into the so called 'Interactive Cave' (Ghali 2003) which was an exhibit that ran in parallel with the event over its final two days. In this cave, instead of guided tours, visitors were invited to roam freely with flashlights while exploring the walls and ceiling to see what could be found. Where flashlights were shone at particular features of the cave, participants would hear ghostly voices narrating stories of its past or, alternatively, extra information relating to the feature in question (see video).

In all, the cave was divided up into three interactive areas in order to cover three distinct walls spanning two linked chambers. Each area was specifically designed to make use of a range of flashlights, playback mechanisms (speakers or headphones) and types of audio content. The targets chosen as content hotspots varied in size and nature, with some, for example, being associated with very obvious feature points whereas others were more subtle in nature and visibility. The three interactive areas (referred to as individual caves for clarity) are described in detail as follows. Pictures and layouts for each cave depicting target, lighting and equipment locations are shown in Figures 2.9 – 2.11

Fig 2.9 – Cave 1: The largest cave, featuring long beam throws requiring a powerful flashlight. A barrier made this cave inaccessible hence increasing the benefit of using flashlights to harmlessly explore it. The barrier was also used as a pivot device by some younger children, as a means to steadily control a heavy flashlight.

Cave 1: In the first of the caves a visitor might enter, three targets were associated with physical objects placed at ground level (two bricks and a fake spider) and a further two were associated with existing holes in the wall. The cave featured a wooden barrier (Fig. 2.9) that defined a natural viewpoint preventing the public from approaching the wall by a distance of approximately five meters. The camera monitoring the space was located beyond this barrier and this meant that there was little chance the visitors could occlude its field of view. Given the long distance to the wall from the barrier, visitors were provided with a relatively large and heavy flashlight (weighing 860g) with which to search the area. The audio content in this cave described its use as a storeroom and larder and this was played through two nearby open speakers aimed into the cave, so as not to be heard too loudly elsewhere.

Fig 2.10 – Cave 2: Users shone flashlights at a set of equally spaced recesses in the wall to trigger content. Headphones were used to trial the medium but also to minimise audio interference between different areas

Cave 2: As well as being a content area in its own right, this cave additionally acted as a thoroughfare between the other two. Headphones were used for audio playback here (Fig. 2.10) not only to minimise aural overlap between the first and last caves (which used speakers) but also to make comparisons between the two playback techniques. In this case, five 40cm targets were associated with man-made recesses in the walls which were approximately 20cm apart and hence much closer together than targets found in the previous cave. Here, given a visitor's close proximity to the wall (max. 2m) the provided flashlight was smaller (weighing 150g) and additionally, trials were carried out in this cave using head-mounted torches to see how these compared to their hand held alternatives.

Fig 2.11 – Cave 3: The monitoring camera was placed to the left of the stairs viewing the active surface at a steep angle. This allowed a wider area to be monitored and also for visitors to stand very close to the wall without causing occlusion. A target located outside the spotlight illuminated area was particularly successful as it needed to be illumined by a flashlight in order to even be seen

Cave 3: In the final cave, five targets were positioned on the opposite wall to Cave 2 and associated with a variety of physical features - four recesses of varying alignment and size (200 to 1500 cm^2) and a piece of carved graffiti found high up near the ceiling (Fig 2.11). Visitors were armed with a medium size flashlight (460g) and sound was again played out through a pair of open speakers. Here, although audio content was themed to tell the story of King David using clips that were mostly less than five seconds long, one clip was deliberately extended over fifteen seconds in order to observe how users responded to it.

2.3.2 Implementation, Deployment and Configuration

To achieve control of sounds in the cave, the technique used to detect and locate flashlight beams was significantly different from that used for the tent and Sandpit. A significant difference between the cave and the StoryTent or Sandpit, was that the monitored surfaces used within the cave remained static. This allowed a form of background estimation to be performed which was not previously possible. To detect flashlights in the cave therefore, a background image of the region to be monitored (lit by normal illumination levels) was first captured. This would be subtracted from all subsequently captured images to leave only the changes between the two. Depending on the beam strength of flashlights used, a threshold was then set against these values, and this would leave only the regions bright enough to contain flashlights as potential candidates. The coordinates of each of these potential candidates were then determined by use of their centre point and these points, when overlapping predefined target areas, would trigger a sound to play.

Deployment of interactive flashlights within the cave was the most challenging of the three installations discussed so far. Access, as expected, was not easy with regard to transport of equipment and the conditions within the cave were also very dark. In particular this made configuration of cameras a significant challenge. It became hard to avoid excessive noise, while maintaining a high enough frame rate

for the system to remain responsive. Positioning was also an issue as the limited space made it hard to place cameras appropriately. In order for the system to be experienced as we desired, these needed to be positioned so it was possible to view all the required targets. Additionally, cameras had to located where they would not be easily noticed or obstructed by visitors. It was noted however that it was not a requirement for cameras to view a surface straight on in order for the system to work. This was exploited to significant gain in the third cave as here, it was possible to position the camera very close to the surface it was monitoring (see Figure 2.12), while still being able to view all the features that we wished to define as targets.

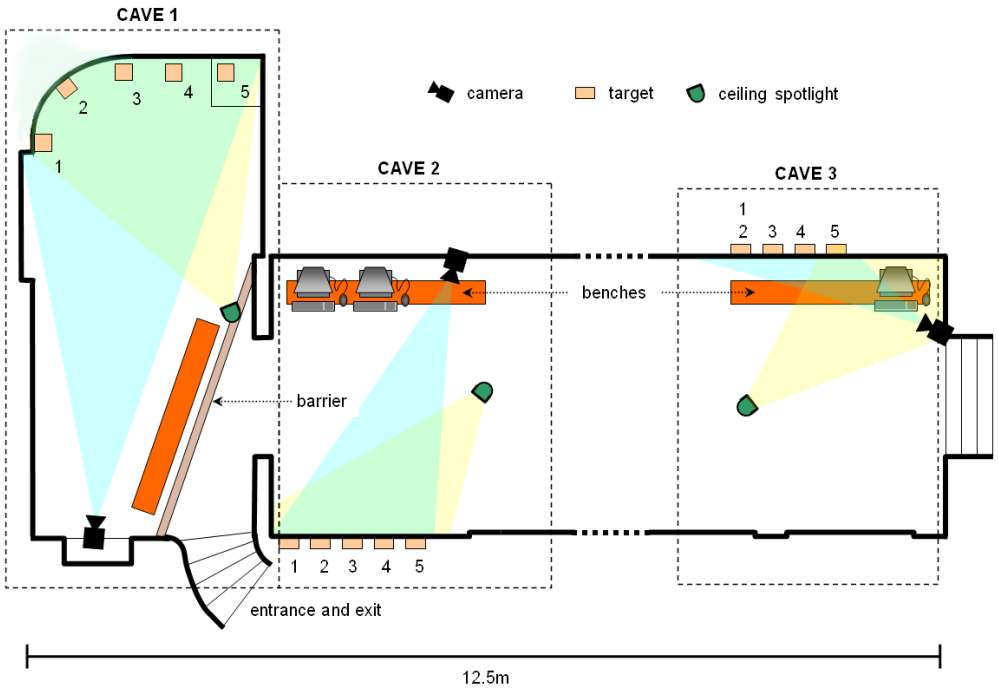


Fig 2.12 – Plan showing the layout of the Caves under Nottingham Castle and how they were subdivided into three areas (Caves) each monitored by a separate camera (fields of view depicted in blue) and with its own set of content targets. Note the close proximity of the camera in Cave 3 to the monitored wall.

To define targets, a 'point and click' interface was employed to place markers against an image of the monitored scene. These could then be sized, moved and have sound clips associated with them using a dialog (Figure 2.13). It was also possible to define an ambient background sound in addition to individual target areas. In the cave, a wind sound would be triggered whenever flashlights were

found to present with the monitored area, and this provided feedback to the users that they were searching within an augmented location.

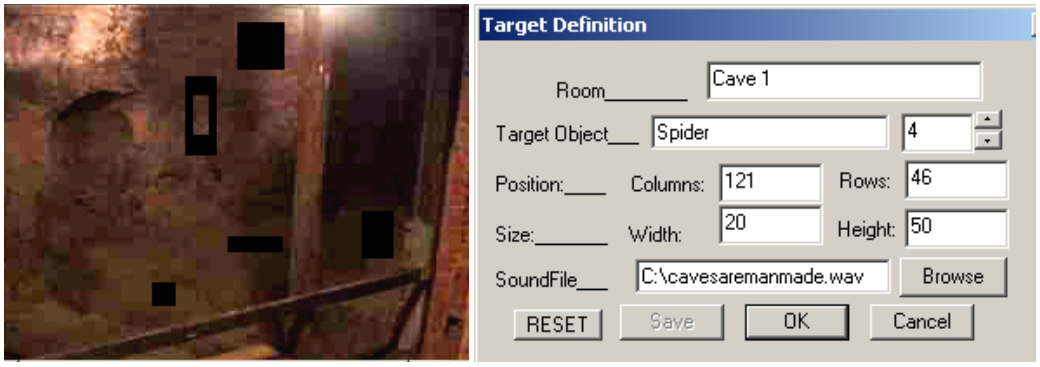


Fig 2.13 – Configuring target positions and their associated sounds via a 'point and click' plus dialog interface

2.3.3 User Trials and Observations

During the two days on which the Interactive Cave was trialled at the Living Exhibition, it was used by over one hundred and fifty visitors aged four years and upwards. Of these, roughly one hundred were adults and fifty were children and the majority of visits involved groups of two or more friends or family members, although there were a few individuals. When visitors entered the cave, as with the StoryTent trials, minimal instructions were given and instead participants were left to explore the cave on their own to see what happened.

From a technical perspective the system was observed to work well. Despite issues where a user unavoidably stood between a camera and the flashlight beam (hence obscuring it), or flat batteries in flashlights caused them to become too dim to trigger targets, no significant breakdowns appeared to occur. Instead, most observations regarding this system were related to its use, and how things might be improved in the future, in order to create a more natural and streamlined experience. One of the most successful elements of the trial, for example, was the fact that the interactive flashlight system made it possible for visitors to point an everyday tool at natural features of a rough cave wall, and have these augmented as a result, without the technology ever becoming apparent. Indeed the target

observed to induce the greatest response from visitors was graffiti carved near the ceiling of Cave 3 which was associated with the legend of King David carving the story of Christ in the walls with his fingernails. This was partly due to the audio content itself tending to promote a strong reaction (e.g. *"uugghhh ... with his fingernails!"*). Also though, because the target was associated with an interesting physical feature that, in fact, couldn't be easily seen without the aid of a flashlight, this made it a particularly satisfying discovery for those who found it.

The fact that it was not always clear which features in a cave were targets, and which were not, tended to encourage two different types of search strategy. The first and most popular of these involved immediately trying out the most obvious physical features then, if successful, applying rules of consistency to find others. A secondary method, commonly employed towards the end of a turn, involved 'painting the wall'; effectively carrying out a systematically exploration. This appeared to be something akin to a last resort, i.e. in order to be absolutely certain no targets had been missed before moving on. However, a problem noted with this technique was that, when 'painting' at speed, a user would often misunderstand the location of a target, commonly believing it to be slightly left or right of its true position, depending on their direction of beam movement at the time. This was caused by a delay between the time a sound was triggered and the point at which a user noticed it so, to combat this, an alteration was made to the system causing sounds to cease playback once a flashlight beam had left their target area. Although this had the effect of largely improving the accuracy with which users could identify targets, it also induced a large change in how they were observed to use the flashlights.

The head torch, for example, was initially expected to be one of the more successful elements of the trial as it had the unique property that, whatever the user looked at was illuminated without the need for hand to eye coordination. Once sound cut offs were implemented however, use of the head torch became difficult since distractions (e.g. checking on children or responding to members of a group) often meant interrupted playback and in practice it became very hard for

visitors to hold their head still for the duration of a clip. The headphones, used in conjunction with the head torch cave (Cave 2), also proved to be detrimental to the experience of users who attended the exhibit in groups.

Similar to the difficulties experienced with head torches was the issue of wobble which has also been observed in the use of laser pointers when interacting with large computer displays (Myers 2002). Here, jitter of the beam resulting from unsteady hands caused problems for fine grained interaction and, with flashlights in the cave, a similar problem caused beams to wander on and off a target. This would result in a sound repeatedly stopping and beginning playback again from the start of the recording which often led to frustration. The problem was particularly challenging for weaker visitors (e.g. young children) when using heavier flashlights over longer distances. This was because, the longer the throw the less angular movement was required in order to move a beam a small distance which was similarly observed in the use of laser pointers. Unfortunately, the longer throws used in the caves required more powerful and heavier flashlights which in turn exacerbated the problem often eliciting giggles from young children hence making their aim even worse.

When applying the work done with laser pointers to the issue, one solution lay in filtering position readings over time so that sudden jumps away and back to a relatively stable position are effectively ignored. With flashlights however it was observed that the issue may have been partially exacerbated by two other effects which were unique to our system. These were:

- Visitors potentially saw flashlights as devices that cast a *pool* of light rather than a single point and thus expected different functionality

- The extent of targets was not always obvious meaning triggers could occur right on their boundaries, hence causing stutter

To explain the former of these, if a user believed *any* illumination of a target was enough to trigger it, this would be in direct conflict with the system design i.e. that the flashlight was effectively some form of degenerate laser pointer. As shown in Figure 2.14, such a conflict could easily lead to confusion regarding triggering and hence exacerbate the wobble problem. A better solution, in this case, might be to utilise overlap between targets and beams and additionally, these proportions of overlap could be used to provide an indication (potentially, for example, via playback volume, see Chapter 5) of when a beam is close to a target boundary. In such a system, it would be important of course that the detected extent of a beam corresponds directly to the user's perception of it (Chapter 5). If this were not so, confusions similar to those observed in the cave may again occur

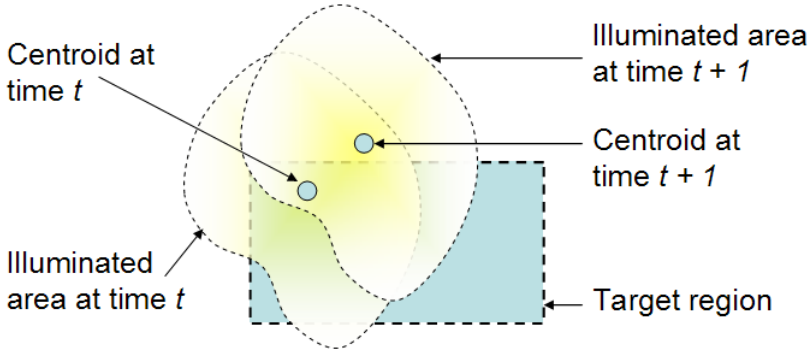


Fig 2.14 – A small change in torch position or orientation can move its centre point outside the target region and cause playback to stop (even though a beam may still be significantly covering a target). This could be avoided by use of overlap as a triggering mechanism

As noted in section 2.3.2, there was a limitation in the cave in that only certain areas could be monitored by cameras and this restricted interactivity to just three regions, the boundaries of which were not visible. The technique used to indicate these areas to the public (by playing the wind sound whenever a flashlight beam was in range) was clearly understood however. Recall that during the course of the trial, users were not explicitly told what to expect. Consequently, there were many examples in which they would explore non-interactive regions of the cave, discover the wind sound and, by determining its boundaries, would go on to trigger audio clips within. There was one exception to this, however, in Cave 3. Here, features of the rock surface formed a visible contour that many users seemed to expect to mark the boundary of the interactive region. This was not the

case. Due to the acute angle that the augmented surface was viewed from, the monitored area did not correspond to this boundary. Since the mismatch was observed to cause some confusion, it would appear that an advantageous extension to the system might be to allow the definition of a interactive area as subset of a camera's field of view. This would not only permit close alignment between those areas perceived to be interactive and those which actually are, but would also make it possible to mask out monitored areas that may be detrimental to system performance. These include commonly occluded regions or areas where shadows and light might be cast.

As a final point, it was interesting to note that, in some cases, users would continue to explore the cave using a flashlight even when they had confirmed there was no more (augmented) content to be found. This shows that it may be beneficial to have an interactive device that is useful in its own right. Such a device can still be used whenever technology is busy (used by others) or even when it fails.

In summary, the main issues discovered during trials with the cave based flashlight interface were related to user interaction. These were:

- A necessity for flashlights to be able to start and stop triggered content in order for users to accurately locate targets

- Difficulties with the steady control and aim of heavy flashlights over long distances (particularly apparent with young children)

- Triggering of targets may not have been in line with user expectations

- Stutter (playback restarts) may occur as a result of target extents being difficult to determine

To conclude, the initial trials with the flashlights as a means to non-invasively augment a static and natural area were successful with a large amount of positive feedback being generated by visitors who participated in the trials.

2.4 Conclusion

With the aim of exploring the feasibility and benefits of a flashlight based user interface, we have looked in detail at the development and deployment of three such interfaces, each utilising flashlights in a varying manner. Although two of the described interfaces make use of similar detection mechanisms (the tent and sandpit), all three have been shown to have separate benefits and issues deriving both from interactive and technical perspectives.

Despite the issues, based on observations and overall user feedback, flashlight interfaces appear to be feasible, both technically and from a usability point of view. The interfaces described here were found to be attractive and enjoyable to use. We conclude from these early installations that flashlight interfaces are worthy of further research.

2.5 Summary

In this chapter our initial attempts at creating interfaces based on visually tracked flashlights are presented. For each interface, its context, construction and deployment is described. Observations are made regarding how successful the trials with each interface were, together with recommendations as to how such interfaces might be improved. Based on these recommendations, in the next chapter we begin to explore possible ways to develop an enhanced flashlight interface, by investigating how flashlight beams can be detected in a manner that is more robust than the techniques so far discussed.

Chapter 3

Visual Detection and Description of Flashlight Projections

The trials described in the previous Chapter with the StoryTent, Sandpit and Nottingham Castle Caves demonstrated that domestic flashlights can form the basis of a usable and engaging vision-based interaction device. These early installations employed standard, rather simplistic, computer vision methods that were adequate but suffered some limitations.

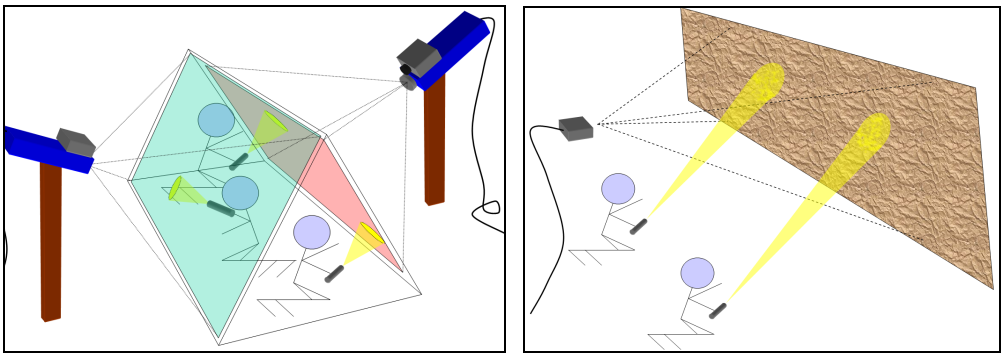
The key requirements of a flashlight-based interactive device are that the vision methods involved be able to reliably detect and describe the projections of users' flashlights onto the target surface. In real interactive installations, detection is complicated by the need to locate flashlight beams over a wide range of surfaces and under a variety of illumination conditions. The ability to differentiate between different beams would allow varying content or effects/controls to be associated with individual flashlights. To achieve this, a description of each flashlight projection is required that is rich enough to support recognition.

The image thresholding approach used to detect flashlights in the StoryTent was sufficient to allow children to interact with the projected virtual environment, but required the flashlight projections to be the brightest objects in view. This is hard to guarantee, and makes system setup difficult. It encourages the use of unnecessarily bright flashlights, leading to high dynamic range images that are difficult to work with. Use of very bright flashlights often results in the corresponding image regions being saturated, reducing the amount of information available for their description.

The background subtraction method used in the Nottingham Caves was effective, but responds to any object, flashlight or otherwise, that is not present in the

background scene. A more flashlight-specific method would increase robustness and should make more useful descriptive information available.

The remainder of this chapter is concerned with methods for the vision-based detection and description of flashlight projections. The installations described in Chapter 2 used different image acquisition scenarios. In the StoryTent (Figure 3.1i) flashlights were usually shone on a translucent screen from inside the tent, while a projector and camera were mounted outside. The flashlight projection was therefore viewed through a screen on which a time-varying virtual world was displayed. In the caves, flashlights were played over physical surfaces, reflecting light into a camera in the normal way (Figure 3.1ii). The Sandpit mixed these configurations, with flashlights and camera aimed towards a physical surface onto which a display was projected.



i) ii)

Fig 3.1 – Flashlights through semi transparent projection surfaces ii) and also reflected off physical surfaces

In what follows these scenarios are considered separately, with the aim of examining the possibility of detecting and describing flashlight projections, in an enhanced manner, in each situation.

3.1 Transmitted Flashlight Projections

In the field of computer vision, detection and tracking is commonly related to observing either rigid or deformable physical objects. Though the determination of illumination conditions has attracted some attention, research on locating light

sources within images of a particular scene (e.g. a light bulb) appears scarce. This may be due to the problem being an uncommon goal or simply because it appears trivial and probably solvable by looking for the brightest point in a scene and thresholding (Chapter 2).

Our aim is to detect the beam of the flashlight that is transmitted through the surface of the StoryTent which, at the same time, features an unconstrained and changing projection incident on the monitored side of the fabric. Under these circumstances, unless the projected scene is carefully designed, there is no guarantee that the brightest point found will always be the result of a flashlight. A detection technique that is more attuned to this specific situation is therefore desirable.

3.1.1 Source Operators

Previous vision research, regarding lighting in a scene, is commonly either concerned with minimising (Stauder 1999, Brelstaff and Blake 1988) or using (Woodman 1980, Barsky and Petrou 2003) illumination effects. The recovery of illumination conditions has also received some attention. Methods have been reported which determine illumination source (Kanbara and Yokoya 2004), the size and location of multiple area light sources (Zhou and Kambhamettu 2004, 2002), dominant illumination direction (Nillius and Eklundh 2001) and distribution (Okatani and Deguchi 2000). Work in this area, however, typically does not expect sources to lie within the field of view. In contrast, Ullman (1976) describes work towards this very goal. Moreover, Ullman's work is set within an environment that appears very similar to that described for our tent scenario.

With an aim to simulate the human ability to detect light computationally, Ullman devised an experiment. If successful, this would eliminate the possibility that high-level analysis and application of cognitive knowledge was a factor in the human ability to perceive light. The results of Ullman's experiment would then be

used to evaluate the contribution (or lack of) of various other likely candidates to the task.

The experiment made use of Achromatic Mondrians (Land and McCann 1971) which constrained its subjects to views made up of different coloured rectangles of overlapping paper. This stopped them applying higher knowledge of light sources to the situation as the viewed scene was simplified by discarding shadings and recognisable sources such as light bulbs (Fig. 3.2). During the course of each session, lights were placed randomly behind different sections of the Mondrian and these were slowly increased in intensity until they became apparent to a human subject. Note was taken of the points at which light sources were *possibly* detectable (subjects were uncertain) and of the points at which they became prominent.

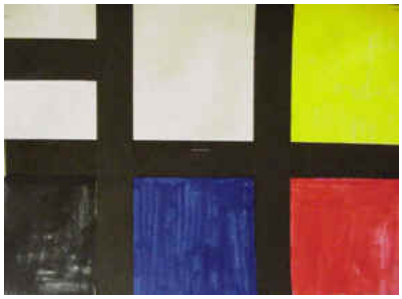


Fig 3.2 – A typical Mondrian

The experiment successfully showed that humans could detect light even in a Mondrian world (which is comparable in appearance to the tent) where it was difficult to apply higher level knowledge to the situation. To discover how this was achieved, six possible classifiers, that could be emulated computationally, were considered (Ullman 1976). None of these however were found to be wholly applicable to the results collected and hence were unlikely to be the dominant factor used by humans to detect light. Instead, Ullman theorised a new method that might be used for light detection in this scenario. Since it is possible to draw comparisons between the Mondrian environment and that of the tent, in that both feature a surface through which transmitted light is to be detected, Ullman's method appears applicable to the tent scenario.

The issue with detecting transmitted light in a scenario such as the Mondrian world is the difficulty of determining whether light measured over a particular region is simply a result of ambient illumination reflected off that surface (Lambert 1760), or if, additionally, the region is also being lit by a light source from behind. Without prior knowledge of the reflectance properties involved, it appears impossible to do this. Ullman's technique (termed the 'Source' or S-Operator) proposes to solve this problem, of determining between reflected light and transmitted light, using only spatial measurements of intensity and intensity gradients. Both of these can be recovered from a standard image capture. The method works by examining two points (*0* and *1*) which are close together, but considered to be 'potentially' either side of a suspected light source boundary. By utilising the aforementioned properties, estimated at each of these points, it is possible to determine the existence of a light source present at one of these locations.

For the method to work, Ullman relies on three key assumptions. These are: that transmitted light through a surface is additive to light that is reflected off the surface, that transmitted light is uniform in intensity and, finally, that the ambient illumination on the scene (I) has a linear gradient (K) over the area between the two considered points. It is noted that the last of these assumptions is only likely to be true if the distance between the points (d) remains very small.

In the described scenario therefore, the ratio of intensity gradients found at each point (S_0/S_1), regardless of their underlying surface reflectance properties (r_0 and r_1), should be equal to the ratio of the two points' measured intensity values (e_0/e_1) if the points are so close that the light illuminating them (I) is approximately equal. If these ratios are not equal, then the discrepancy is caused by the presence of a light source (L) at one of these points. Because such a light source is considered to have a uniformly additive effect, its presence will increase the measured intensity at this point, without changing its intensity gradient.

Ullman's S-Operator is derived from Lambert's equation where θ is the angle between the point of observation and the surface normal

$$e = Ir\cos\theta \tag{3.1}$$

which is simplified and altered to include the additive effect (which may be zero) of a transmitted light source (L) at point *0*

$$e_0 = L + I_0 r_0 \tag{3.2}$$

This is rearranged to give

$$L = e_0 - I_0 r_0 \tag{3.3}$$

Similarly we find at point *1*

$$e_1 = I_1 r_1 \tag{3.4}$$

By assuming ambient illumination at point *1* (I_1) can be represented using that at point *0* (I_0), together with the ambient illumination gradient (K) and the distance between points (d)

$$I_1 = I_0 + Kd \tag{3.5}$$

I_0 can be replaced in eqn. 3.2 which can then be expanded to form

$$L = e_0 - I_1 r_0 + r_0 Kd \tag{3.6}$$

Here, since the multiplied (and unknown) values r_0 and K are equal to the intensity gradient (S_0) at point *0*

$$S_0 = r_0 K \tag{3.7}$$

these can be replaced, as can I_1 (by a rearrangement of eqn. 3.4) to form

$$L = e_0 - e_1\left(\frac{r_0}{r_1}\right) + S_0 d \tag{3.8}$$

Since a substitution and rearrangement of eqn. 3.7 into

$$S_1 = r_1 K \tag{3.9}$$

leaves

$$\frac{r_0}{r_1} = \frac{S_0}{S_1} \tag{3.10}$$

the unknown surface reflectance ratio, between the two points, can be replaced with the known ratio of their intensity gradients. This gives the final S-Operator

$$L = e_0 - e_1\left(\frac{S_0}{S_1}\right) + S_0 d \tag{3.11}$$

Having examined the technique in detail, it is clear that such a method cannot be applied to the tent scenario. This is, in part, due to conflicts in its environmental assumptions with the StoryTent environment but also due to impracticalities of its actual implementation.

As can be seen in the S-Operator's derivation above, the requirement that ambient illumination must vary in a linear fashion between input points is absolutely core to the technique. This simply cannot be said to be true in the tent scenario. Although the different coloured regions of the tent projection might appear synonymous to 'Mondrian-like' areas of varying reflectance, featured in Ullman's experiment, they are not. They are in fact variations in ambient illumination (*I*)

which are unconstrained and liable to change in a non-linear fashion. It is folly therefore, to assume they will exhibit a linear gradient (*K*) between considered input points.

In addition to conflicts in environmental assumptions with the tent scenario, the S-Operator also suffers a number of issues with regards to implementation. Since the operator's input points must lie either side of a light source boundary, in order for its presence to be detected, if the points are close together (as is required in order for the linearity assumption to remain true), the technique will only ever mark a 'light source border' (rather than region). In fact, only one side of the region will be marked (corresponding to the scan direction). To determine the extent of an entire region containing a light source therefore, a method of grouping results from multidirectional scans would be required (e.g. Fig. 3.3)

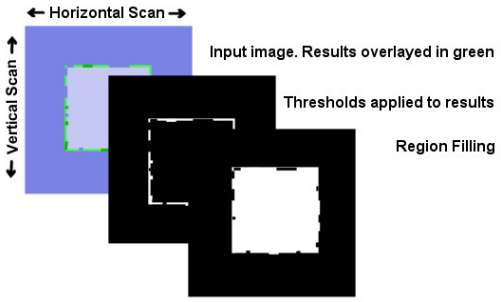
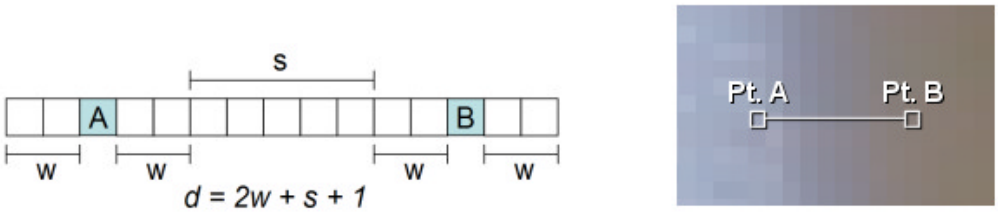


Fig 3.3 – Potential (additional) stages of processing required in order to use S-Operators to mark an entire region containing a light source (rather than the side determined by the operator's scan direction).

Similar issues arise, with regard to input point separation, when considering that an estimation of each point's illumination gradient is required. For accuracy, it is best practice to measure such a gradient over a few pixels either side of the point in question (Fig. 3.4i). Since these 'gradient neighbourhoods' cannot intersect with one another, this forces further separation between input points thus making the existence of a linear illumination gradient between them less likely. This situation is further compounded by the need to be able to measure either side of a step edge (which can be quite wide: Fig. 3.4ii). Additionally, if the gradient neighbourhoods used at either of the two input points incorporate such step edges, the gradient at this point will be falsely estimated. This can lead to erroneously strong operator responses.



i) ii)

Fig 3.4 – i) To determine accurate measurements of intensity gradients at each point, values must be taken from within a small 'gradient neighborhood' (*w*) either side of each point. Since such neighborhoods may not overlap, the distance between input points (*d*) must accommodate them. ii) To obtain good results the distance between input points must be large enough to span the width of step edges.

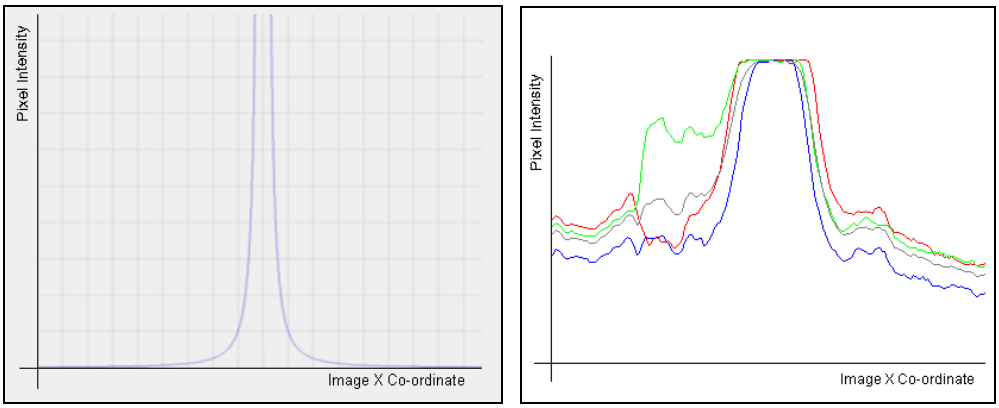
Although the projected lighting, found in the tent scenario, is enough to rule out use of S-Operators in this environment, the above implementation impracticalities are also worth noting. These highlight potential weaknesses with the technique that may present problems even when S-Operators are used within environments that conform exactly to Ullman's described assumptions.

3.1.2 Specularities

A second possibility for improved flashlight detection in the StoryTent, or indeed the Sandpit scenario, lies in the consideration that, in these situations, flashlights commonly appear as small bright regions, a description that can also be applied to specularities. Specularities are defined by Brelstaff and Blake (1988) as, *"bright image regions formed by specular reflection"* which, *"occur whenever a glossy surface reflects light in a mirror like fashion"*. Specularities are therefore points at which a light source is effectively shining directly into the camera, as is the case in StoryTent.

The profile across a typical specularity consists of a sharp increase in intensity, often to the maximum brightness detectable, followed briefly by an identical drop (Fig. 3.5i). In comparison, figure 3.5ii shows a similar profile, in this case taken across a flashlight as viewed on the surface of the tent. Here, although the intensity does not exhibit as sharp an increase as would be present over a real specularity, the similarities in shape are still apparent. Despite over exposure of

flashlight beams in these scenarios being common (as shown in figure 3.5ii, this creates a flat peak in an intensity profile) the model of *"a small region rising sharply in intensity to a single point"* seems generally applicable. A detector capable of identifying specularities (Brelstaff and Blake 1988, Forbus 1977) might therefore be adapted to detect flashlights in images captured from the StoryTent and possibly even the Sandpit scenario.



i) ii)

Fig 3.5 – A typical specularity intensity profile, i) compared to that of a flashlight, ii). In ii), standard colours represent RGB components and grey, the overall intensity.

Use of specularity detection techniques for flashlight detection, while potentially improving system stability, does not however provide extra opportunities to significantly enhance the user experience of such installations as the StoryTent and Sandpit. As figure 3.5ii shows, flashlight projections typically appear as saturated regions of the input image. Any information regarding the pattern of light projected by a given flashlight is therefore lost, and only location and crude shape information remains. As flashlight projections are all similarly shaped, the development of flashlight recognition, for example, becomes almost impossible.

While improvements to flashlight detection, might be possible in scenarios that incorporate projections (for example using Brelstaff and Blake's Cylinder Test, 1988), there is more potential for obtaining rich descriptions of flashlights in our second scenario (Fig. 3.1ii). For this reason, the work reported in the remainder of this chapter focuses on the description of flashlight projections, given images of their reflections off an unaltered physical surface.

3.2 Reflected Flashlight Projections

In the previous section we considered how flashlight projections might be detected primarily because, in the StoryTent and Sandpit scenarios, it was necessary to be able to differentiate such projections from other bright regions typically found in a scene. As flashlight projections are usually saturated in the StoryTent and Sandpit installations no information can be obtained about the pattern of illumination being projected by the flashlight.

When flashlights are reflected from an unaltered physical surface (Fig. 3.1ii), information about the illumination patterns they emit is apparent in the images obtained. If an accurate description of such profiles could be obtained, this could not only be used as a means to recognise individual flashlights but also to distinguish them from other objects (e.g. people).

3.2.1 Recovering Flashlight Illumination Profiles

One method, which might be utilised to obtain such a description, lies in the work carried out by Tsukiyama (1995) which analyses the illumination from a point light source on a scene. This is done in order to infer the 3D shape of plane surfaces that might be present, specifically, to recover such surfaces' relative orientations. Tsukiyama supposes that, when surfaces are illuminated by a single or dominant point light source, the location of each surface's specularity marks the position where the surface is exactly perpendicular to the light source in question. When considering two surfaces which are positioned at differing angles to such a source, these specularities would be located at separate positions. Knowledge of these positions would allow the angle between the surfaces to be calculated as well as its nature (convex or concave). This would be done using the order of the points, their separation from one another and the perpendicular distance of each surface from the light source.

The challenge in this technique lies in the fact that specularities are not always present. This can occur because surface properties, or diffuse light, prevent them from being formed or, alternatively, because the position where a surface's specularity *should* be formed, lies beyond its extent, or outside the image. Since the presence of visible specularities cannot be guaranteed, analysis of isoluminance curves, present on each surface, is instead used to determine their centre. This indicates where specularities would be located if they existed. The technique used to perform such analysis, might also be used to obtain descriptions of flashlight illumination profiles.

The method works by subtracting the edges found by Sobel edge detection (Sonka, Hlavac and Boyle) and grouping the remaining pixels by which surface they belong to. This is done using a region labelling operator discounting any areas where the number of enclosed pixels is below a set threshold. Using such a restriction ensures only true surfaces are extracted. The pixels in each surface region are then grouped into segments of approximately equal intensity (0.5 - 2 grey levels) and these form a rough estimation of the isoluminance curves attributable to the light source. An attempt is then made to fit circles to each of these curves (using a least squares metric) and potential centre points are plotted into a 2D accumulator array. Ignoring circles with excessively large or small radii, the centre of irradiation is determined by selecting the central coordinate of the most popular cluster in this array.

Despite having a sound theoretical basis, the technique suffers from a number of inaccuracies. When used for its original intended task, Tsukiyama notes that results are affected not only by additional light sources (that significantly change scene illumination), but also by concave surfaces reflecting light onto one another (essentially emulating an entirely new, if weak, light source). The approximation of fitting circles to isoluminance curves (rather than ellipses) also introduces imprecision.

When considering how applicable the method is to describing flashlight projections, not all of the above issues are worthy of consideration. There is however, one very significant flaw in using such a technique. This is that the reflectance across each surface must remain uniform, or it will severely affect the accurate construction of each profile's isoluminance curves. When working with naturally plain surfaces of course (as Tsukiyama does), this does not cause a problem. However, any surfaces containing content a user may wish to illuminate will naturally contain frequent changes in their reflectance properties. Such changes will often occur over areas too small to be considered a separate region in the same plane. In these cases, not enough data would be available to obtain accurate curve estimations. The technique is therefore likely to break down.

In order to exploit this, or in fact any method of describing flashlight projections, a means to remove the effects of varying surface reflectance on results, is required. For further development of a flashlight interface it is desirable, in fact, to be able to find the flashlight intensity incident on a surface irrespective of *any* altering factors at all. To do this we must first consider what these factors are, and then examine ways in which they might be removed.

3.2.2. Flashlight Projections, Ambient Illumination and Reflectance

A common vision technique for isolating moving objects against a static background is to subtract the background from a foreground image, using absolute differences, then threshold the result to create a mask. This mask is then applied to the original foreground image to make only the moving region available for further processing (recognition etc).

Fig 3.6 – Results of using to common background subtraction to isolate opaque objects in images

When used with opaque objects (Fig. 3.6), such a methodology is valid as the items in question obscure the background they are positioned over. Since reflected flashlight beams are not opaque however, application of this technique to our scenario would provide regions which do not represent just the illumination from the flashlight (I_t – t refers to the synonym 'torch' to mean flashlight) but additionally the effect of the surface's reflectance properties (R) together with that of any ambient illumination (I_a). Specifically, by applying a simplification of Lamberts equation, the image data retrieved from the isolated regions in our foreground image (e_f) would represent

$$e_f = (I_a + I_t)R \tag{3.12}$$

assuming the light from both the flashlight and the ambient illumination is additive. Rearrangement gives

$$I_t = \frac{e_f}{R} - I_a \tag{3.13}$$

which leaves the unsatisfactory situation of having two unknown values, I_a and R contributing to results.

A technique to remove the first of these unknown's contribution to our results lies in the method employed during the Caves event (Chapter 2) i.e. to simply subtract the foreground image (eqn. 3.12) from the background (e_b)

$$e_b = I_a R \tag{3.14}$$

to give

$$e_f - e_b = (I_a + I_t)R - I_a R \tag{3.15}$$

which can be expanded and rearranged to form

$$\frac{e_f - e_b}{R} = I_t \tag{3.16}$$

As can be seen from eqn. 3.16, although ambient illumination is no longer a contributing factor, in order to achieve the goal of extracting illumination from flashlights on its own, the results of subtracting one image's intensity values from another's must additionally be factored by the reciprocal of the reflectance. Since it is impossible to discover the reflectance properties of a surface at a particular point, without first knowing the illumination at that point, such factoring cannot be carried out. Use of this technique would therefore leave results that are factored by the unknown surface reflectance properties of each point in the image.

Using extremely controlled conditions, it is possible that the technique could be applied under the assumption that the surface being monitored is of uniform reflectance. Although recovered data would not be an exact representation of the illumination emitted by a flashlight, because, in this case, all values would be factored by the same amount, the illumination pattern cast by the flashlight would remain constant no matter where in the image it was directed. This would allow recognition for example to be carried out. Unfortunately, such an assumption would place unacceptable constraints upon the usefulness of any given system. It seems unlikely therefore that this technique represents the best solution.

An obvious alternative here is to base results on values that are affected by ambient illumination on a scene instead of surface reflectance. Although, between acquisition of fore and background images, ambient illumination is more likely to

vary than surface reflectance, it is an acceptable assumption (at least in the cave environment) that such variations will be sufficiently small as to be insignificant. Ambient illumination therefore can be considered to be an unknown but constant value in both eqn. 3.12 and 3.14. Since reflectance (between image acquisitions) can also be considered constant it can be substituted in eqn. 3.12 for a rearrangement of eqn. 3.14 to give

$$e_f = (I_a + I_t)\frac{e_b}{I_a} \tag{3.17}$$

which when expanded gives

$$e_f = e_b + \frac{I_t e_b}{I_a} \tag{3.18}$$

Dividing through by e_b and re-arranging leaves

$$\frac{e_f}{e_b} - 1 = \frac{I_t}{I_a} \tag{3.19}$$

which can be further re-arranged to give the flashlight illumination as follows

$$\left(\frac{e_f}{e_b} - 1\right) I_a = I_t \tag{3.20}$$

As noted, when using eqn. 3.16, to determine flashlight intensity values, it is impossible to estimate, or in fact assume anything about, the potential variations in surface reflectance over a given set of points. Use of ambient illumination as the unknown variable however (eqn. 3.20), has the significant advantage that, in most cases, the changes in I_a over an equivalent set of points, will either be negligible, or at the very least approximately uniform. By removing or replacing I_a with a constant in eqn. 3.20 (necessary because it is unknown), this calculation,

in theory, should give a spatially correct approximation of a flashlight's profile at any given position in a scene, even if this is not correct in terms of magnitude.

The assumption that ambient illumination incident upon a scene is approximately uniform, is fair for the majority of scenarios. If an illumination gradient is exhibited however (as in one area of the Caves for example), this is unlikely to have a significant effect unless the gradient present is unusually steep. This is because, if the gradient is gradual, the change over the area covered by a flashlight projection will be negligible. It is possible however, to factor out the unknown environment variables I_a and R from the equation entirely, by considering data that is not only obtained from a pair of fore and background images, but also from a second foreground image, taken from a temporally adjacent frame at time, $t+1$.

At any moment, spatially equivalent pixels will have equal reflectance properties and virtually equal (providing Δt is small) ambient light illuminating them. This means the only factor that is likely to change significantly between frames is I_t. As a result, it is possible to re-arrange eqn. 3.16 (eqn. 3.20 could be used also) as

$$\frac{e_f - e_b}{I_t} = R \tag{3.21}$$

then set it equal to itself at the next frame increment

$$\frac{e_f - e_b}{I_t} = \frac{e_{f1} - e_{b1}}{I_{t1}} \tag{3.22}$$

Further re-arrangement gives

$$\frac{e_f - e_b}{e_{f1} - e_{b1}} = \frac{I_t}{I_{t1}} \tag{3.23}$$

Unfortunately, although eliminating both the environmental elements, eqn. 3.23 leaves yet another factor this time manifest in the unknown illumination of the flashlight at time $t+1$ (I_{t1}). The result of this is that, at any moment, it is possible to calculate the intensity of a flashlight projection, irrespective of environment factors, but only as a ratio of its intensity (at a given point) over time. As the flashlight is likely to be in (unpredictable) motion, no strong assumptions can be made regarding this value. Additionally, if the flashlight is motionless between time t and $t+1$, this ratio will become 1 and hence the operator will provide no useful information.

In summary, it appears that it is impossible to recover flashlight intensity incident on a surface, irrespective of altering factors. In each considered option, we are always left with at least one unknown value. Considering therefore, only the calculations (or operators) that feature a single unknown, experimental analysis of each against real data is required to determine the best candidate for the task. The three candidates shall be termed the Subtraction (eqn. 3.16), Quotient (eqn. 3.20) and Ratio (eqn. 3.23) Operators.

3.3 Subtraction, Quotient and Ratio Operators Compared

Section 3.2.2 proposes three possible methods for recovering the intensity of flashlights incident on a surface with minimal influence from external factors. Since it is impossible to remove such factors entirely (without extensive knowledge of the environment a system is deployed in) each operator's results will be influenced by one unknown. These are: surface reflectance properties, ambient illumination levels and a flashlight beam's speed of movement. Based on discussion of the benefits and issues of each of these, the best results, against a typical varying background reflectance, should be achievable using the proposed Quotient Operator.

The remainder of this chapter describes an experimental evaluation of the ability of the Quotient operator to provide accurate and consistent descriptions of the pattern of illumination projected by a flashlight.

Although the Ratio Operator represents the only method of the three that, in theory, produces results that are entirely independent of environmental effects, it cannot actually be considered a candidate for this task. This is because, unlike the other operators, it can never provide flashlight profiles, from which recognition might reliably be achieved, unless the flashlight in question is moving at a constant rate. This is an unrealistic constraint. The Ratio operator will not be considered further.

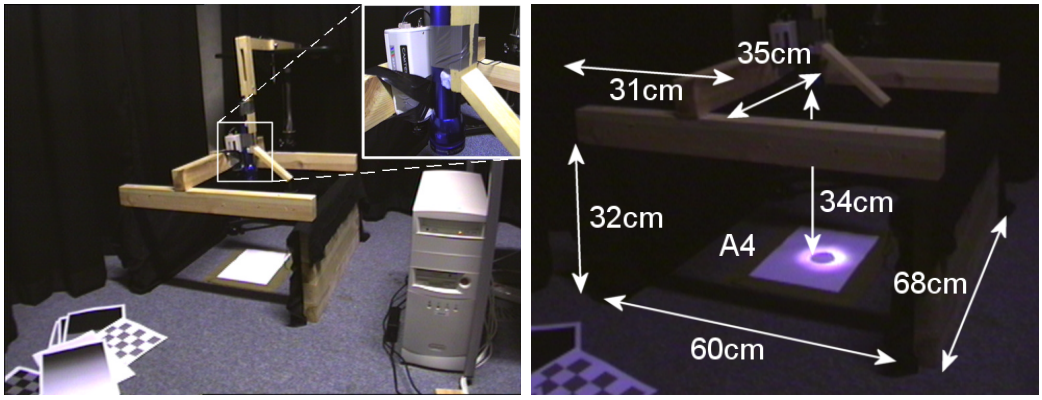
Though the output of the Subtraction operator is a function of both flashlight illumination and surface reflectance, it is computationally efficient compared to the Quotient operator and was used to good effect in the Nottingham Caves. Rastering a division calculation across image pairs creates a high computational overhead. Although the Quotient operator is likely to produce results that are less detrimentally affected by reflectance than the Subtraction Operator's, such results will manifest themselves over a very condensed range. This leaves potential for inaccuracies of measurement or noise to become more apparent in operator output. In what follows the Subtraction operator therefore provides a benchmark against which the proposed Quotient operator is compared.

3.3.1 Design

To perform a fair evaluation of the proposed Quotient operator we create a controlled simulation of conditions that are as close as possible to those under which a flashlight interface might be used. For a typical scenario then, a flashlight is likely to be moving over a surface of varying reflectance, at an approximately fixed distance. Given that such an interface is more commonly

likely to be housed indoors, minimal or zero change in ambient illumination is an acceptable assumption.

To simulate this, a rig was constructed whereby a camera could be positioned alongside and approximately parallel to a flashlight. This was set at a fixed distance above a (replaceable) reflective surface, in order to monitor the pattern of illumination incident on it (Fig. 3.7i). Assuming the operator in question performs as expected, it should be possible to recover profiles of the illumination due to the flashlight in the scene, which are identical, or near identical, regardless of the reflectance properties of the surface. A selection of ten surfaces were produced using a laser printer, each featuring a different reflectance pattern. For each one used, image pairs were taken of it, in a set position. These represented background images, in which the scene was lit only by ambient illumination, and foreground images in which the same scene was additionally illuminated by the flashlight. Using this data, the Subtraction and Quotient operators were applied to each image pair in turn. Results were compared statistically, over the entire set, to assess their ability to produce a consistent representation of the flashlight illumination profile.



i) ii)

Fig 3.7 – Experimental setup depicting rig and camera/flashlight mount i) as well as dimensions ii)

To minimise variation, the experiment was conducted in a lab lit only by normal levels of constant artificial light. Given the situation noted by Tsukiyama, whereby surface reflectance can act as a secondary light source (Section 3.2.1), the rig sides were also positioned a significant distance from the edge of the

camera's field of view (Fig. 3.7ii). These were covered in low reflectance dark cloth to reduce the amount of light incident on the monitored surface, which was not due directly to the flashlight beam or ambient illumination. In addition to this care was required over a number of other factors that would otherwise affect the experiment. Under normal conditions for example, a flashlight in a scene would only be used to illuminate a small portion of the frame in question, the remaining space being used to encompass as much content, relevant to the interface application, as possible (see Chapter 2). For experimental purposes however, we are only concerned with the consistency of the descriptions produced. It is desirable therefore, not only to have as high a resolution image of the flashlight beam's profile as possible, but also that the profile is not uniform in intensity. Such a profile should instead exhibit a rich structure containing clearly visible dark and light regions. If, by applying either operator, such a structure can accurately propagate through to results, the operator will be considered successful.

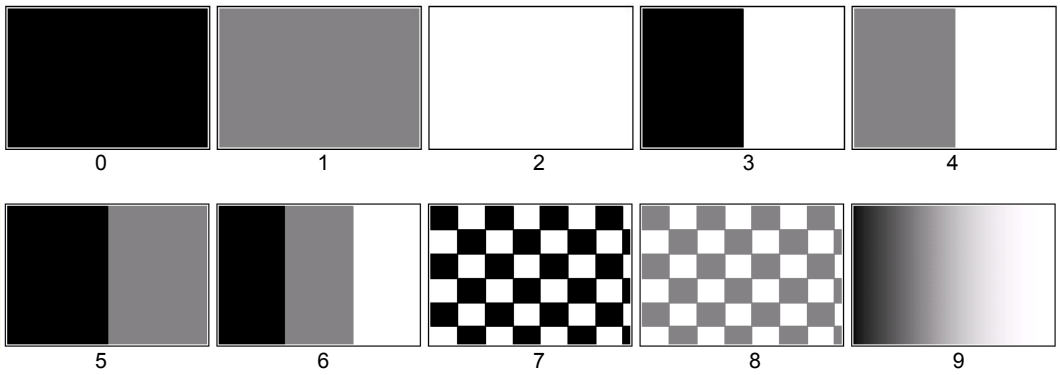


Fig 3.8 – Laser printed reflectance patterns. 0 – 2: minimal to maximum uniform reflectance, 3 – 6: medium to high contrast changes, 7 – 8: frequent medium and high contrast changes. 9: low contrast reflectance variation.

The problem of dynamic range raised in Chapter 2 again proved to be an issue. In order to test the operator's resilience to changes in surface reflectance, it was important that the reflectance patterns produced represented a logical progression of gradually increasing contrasts and spatial extents (Fig. 3.8). This made it possible to determine over what range results could be expected to be accurate. The issue with this is that, as illustrated in figure 3.9, when configuring the camera to monitor a white surface, so that the resultant images actually appear white, any addition of increased intensity would not be apparent. This means that

analysis of the flashlight profile's illumination pattern would be impossible. To combat this, the solution is of course to alter factors affecting the brightness of the image. The large contrasts existing in some of the reflectance patterns however, made this problematic. It was difficult to find a configuration where, for all surfaces, the brightest regions did not saturate when illuminated by a flashlight but also, where regions of low reflectance were still visible under just ambient illumination.

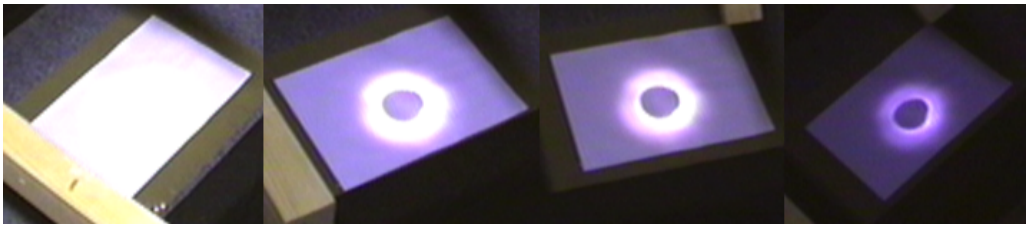
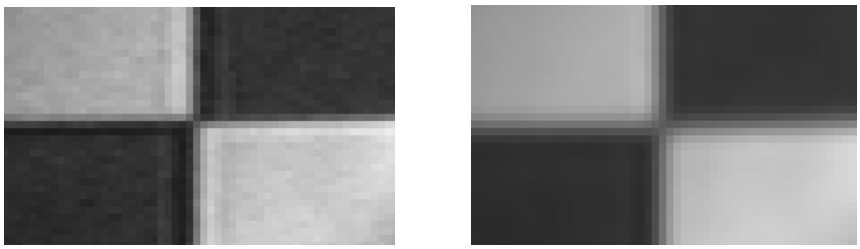


Fig 3.9 – Sequence depicting the required reduction of light entering the camera (captured from a 2nd viewpoint). Left: normal settings required for a white sheet to appear as such. Right: amount of light reduction required before the profile of additional light, directed at the same sheet, can be viewed without saturation occurring.

To achieve the above described requirements, a combination of a low powered flashlight and a reduction in the amount of light entering the camera was used. Although such a configuration allowed operators to be tested against a wide range of contrasts, capturing background representations with only minimal illumination entering the camera introduced a greater potential for noise. To combat this, and to deal with resolution issues regarding errors sensed over step edges (Fig. 3.10), the operators were additionally applied to images smoothed by varying amounts, to see if such filtering had any significant effect on results.



i) i)

Fig 3.10 – Smoothing to counteract noise and anomalies found to occur near step edges in low intensity images

3.3.2 Analysis

The data gained from the described experimentation comprises ten image-sized arrays of output values per operator, one for each of the reflectance patterns employed. These represent the profile of flashlight illumination recovered by the two operators, from each of the ten scenes. Ideally, for each of the two sets of result images, all ten results should appear roughly identical. To evaluate the performance of the operators a number of techniques have been considered.

Ideally, the output of each operator would be compared with a predetermined ground truth which describes the actual pattern of illumination emitted by the flashlight. Errors would be computed for each operator, and those errors compared and contrasted. Unfortunately, no such ground truth exists.

An alternative approach would be to assess the consistency of the descriptions produced. A truly successful operator would generate a set of ten identical output arrays, their contents being independent of the reflectance patterns on the target surface. Mean and standard deviations could be computed from the ten outputs achieved at each pixel location, producing two image-registered arrays of variation estimates, one for each operator. This approach, however, is also problematic. Though it would provide numerical estimates of the variation present in each operator's output, the two data sets cannot be compared. The Subtraction and Quotient operators produce output (and so mean and variance) values in widely different ranges.

To avoid these problems the two operators were instead applied to only nine of the reflectance patterns' fore and background image pairs. This produced nine result arrays as described above. Using each of these result arrays in turn, it was then possible to simulate flashlight illumination on a scene that featured the 10^{th} reflectance pattern. By inverting each of the two operators, each operator's set of result arrays were used in turn with the tenth scene's captured background image, to create a set of synthetic representations of the tenth foreground image. This is

achievable because flashlight illumination (I_t) cast on each scene remained constant throughout. It is therefore possible to rearrange the equations of each operator (3.16 and 3.20) to use their previous results as known values. Such values, in combination with new background image data can be used to estimate foreground image pixel data as shown in equations 3.24 and 3.25.

$$e_F = I_T R + e_B \tag{3.24}$$

$$e_F = \left(\frac{I_T}{I_A} + 1 \right) e_B \tag{3.25}$$

Assuming each operator's previous results are accurate, the set of synthetic images produced from such results should be very similar to the scene's original captured foreground image. Because such images are manifest within the same data range (0-255 levels), they can be directly compared irrespective of which operator's results were used to produce them. By examining the differences between each synthetic foreground image and the original captured one, it is possible to evaluate and compare the stability of the operators.

To apply this analysis technique, it was important to correctly select which of the reflectance patterns available would be used as test pattern and which single one would be used to generate our comparable results. As shown in figure 3.8 (0 - 8), the nine reflectance patterns, chosen for the initial calculations of operator results (test patterns), were designed to progressively increase both the contrasts present and spatial variations of the scene. These start with uniform reflectance at different levels, and then introduce one single step edge, followed by several variations in contrast (grids). Such contrast variations occur over different reflectance levels and magnitudes of change.

The reflectance pattern used to produce the discussed synthetically illuminated images (our comparable results) deliberately contained no such obvious features. It instead represents a smooth alteration in reflectance from the minimum to

maximum levels achievable (Fig. 3.8 - 9). This is because, although variations in reflectance remain beneficial for this pattern, we are interested to see what features from our test patterns stand out, rather than any which occur as a consequence of the one used to generate results. It was important however that the pattern used to generate results, bore no resemblance at all to any from the test set. If such resemblance did occur, then the results obtained using that test pattern, would be mathematically analogous to the subtraction then addition of the same number to a constant. Consequently, uncharacteristically good results would occur in this case.

3.3.3 Results and Discussion

The experimental method employed here generates comparisons between real images of flashlight projections and corresponding synthetic images created by combining a background image (without a visible flashlight projection) with the output of one or other of the operators described in section 3.2.2. Each of two operators was applied to nine image pairs and the results combined with a tenth background image to produce a set of nine synthetic images for each operator. Subtraction of each of these 18 synthetic images from the corresponding real image produces 18 difference images. To summarise this data for analysis the mean and standard deviation of each difference image was computed to give the mean error and the standard deviation of the error. These values are considered here.

Single Background Image – No Filtering				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.830	21.865	14.738	17.874
1	7.829	9.949	5.299	5.294
2	9.290	12.418	3.905	3.824
3	9.942	15.409	7.621	12.199
4	6.419	8.273	3.846	3.906
5	11.714	15.793	8.722	10.920
6	8.156	11.134	6.086	6.750
7	13.070	18.931	8.952	14.632
8	9.658	12.478	4.582	4.768

Table 3.1

Table 3.1 shows the basic data, obtained as described above. Note that in each case, as expected, the Quotient operator provides synthetic images (and therefore descriptions of the flashlight projection) with lower average errors and errors that are more tightly clustered around these lower values. In all but one case the mean error for the Quotient method is less than ten grey levels; the appearance of the flashlight projection on a new background can be closely approximated, regardless of the reflectance properties of the surface from which the flashlight projection was obtained.

Though individual captured images are often used to model the background scene in background subtraction and similar algorithms, it is well known that a single image can be subject to significant amounts of noise. A wide variety of background estimation techniques have been developed (see Chapter 5) to deal with this, and a detailed comparison of these methods is beyond the scope of the current study. To provide an initial assessment of the benefits to be gained by employing an alternative background image creation method that is less sensitive to noise, Table 3.2 shows data similar to that given in Table 3.1, but obtained using the average of three captured frames to form the background image.

Averaged Background Image – No Filtering				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.794	21.833	14.138	17.764
1	7.620	9.945	4.565	4.685
2	9.075	12.449	3.368	3.363
3	9.697	15.461	7.031	12.134
4	6.143	8.277	3.148	3.296
5	11.809	15.869	8.821	11.028
6	7.895	11.149	5.545	6.348
7	12.911	18.955	8.285	14.521
8	9.449	12.515	3.893	4.164

Table 3.2

Once again the Quotient operator provides both lower mean and standard deviation of errors than the Subtraction operator in every case and all are more tightly focussed. Note also that, with the exception of reflectance pattern five, mean errors are slightly lower when an averaged background is used.

Another common response to image noise is to smooth the images concerned with a Gaussian kernel of appropriate size (Trucco and Verri 1998). Tables 3.3 and 3.4 show results obtained when a 3x3 approximation to a Gaussian was used to smooth the foreground and background images. The Quotient operator again produces consistently lower errors than the Subtraction operator, and all mean errors are lower than the corresponding values obtained without image smoothing. Tables 3.5 and 3.6 show similar data achieved after applying a 5x5 pixel approximation to a Gaussian, while Tables 3.7 and 3.8 show the result of applying a 7x7 Gaussian mask. A visual comparison of all these results is provided by figure 3.11.

Single Background Image – 3x3 Gaussian Filtering				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.688	21.860	13.824	18.054
1	7.294	9.963	3.988	4.528
2	8.598	12.507	2.765	3.068
3	9.221	15.547	6.409	12.274
4	5.596	8.267	2.373	2.942
5	11.356	15.826	7.791	10.834
6	7.418	11.259	5.071	6.403
7	12.546	18.909	7.755	14.559
8	9.015	12.450	3.270	3.977

Table 3.3

Averaged Background Image – 3x3 Gaussian Filtering				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.715	21.812	13.371	17.924
1	7.217	9.958	3.673	4.277
2	8.527	12.535	2.568	2.961
3	9.134	15.578	6.142	12.202
4	5.480	8.289	2.101	2.752
5	11.575	15.858	8.050	10.957
6	7.314	11.262	4.819	6.197
7	12.500	18.912	7.397	14.431
8	8.946	12.480	2.950	3.716

Table 3.4

Single Background Image – 5x5 Gaussian Filter				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.682	21.842	13.680	18.074
1	7.204	9.964	3.778	4.420
2	8.466	12.533	2.557	2.955
3	9.070	15.564	6.168	12.280
4	5.426	8.262	2.077	2.772
5	11.306	15.822	7.647	10.817
6	7.268	11.292	4.896	6.359
7	12.396	18.836	7.494	14.479
8	8.868	12.399	3.013	3.837

Table 3.5

Averaged Background Image – 5x5 Gaussian Filter				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.706	21.798	13.255	17.939
1	7.150	9.957	3.516	4.219
2	8.425	12.555	2.423	2.910
3	9.006	15.589	5.965	12.193
4	5.344	8.282	1.888	2.652
5	11.547	15.845	7.939	10.937
6	7.190	11.290	4.692	6.173
7	12.361	18.837	7.188	14.343
8	8.825	12.425	2.753	3.623

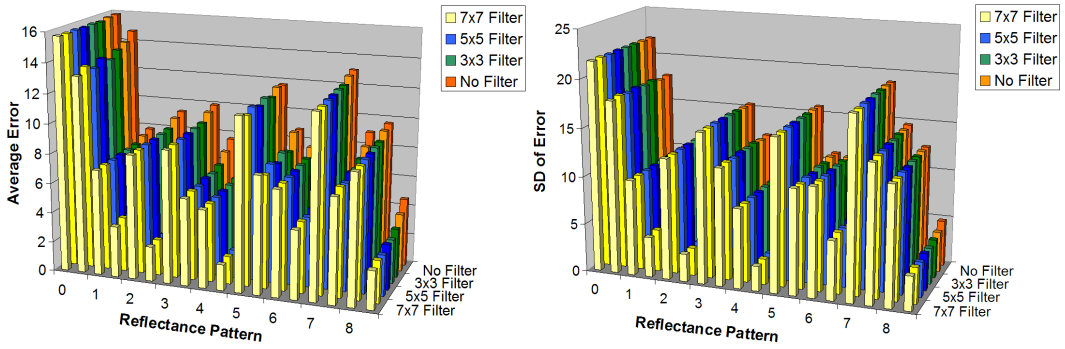
Table 3.6

Single Background Image – 7x7 Gaussian Filter				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.682	21.820	13.599	18.086
1	7.146	9.962	3.649	4.361
2	8.375	12.548	2.415	2.886
3	8.951	15.556	5.999	12.270
4	5.297	8.243	1.866	2.652
5	11.275	15.810	7.558	10.805
6	7.157	11.312	4.783	6.330
7	12.235	18.727	7.289	14.371
8	8.734	12.328	2.827	3.743

Table 3.7

Averaged Background Image – 7x7 Gaussian Filter				
Reflectance Pattern No.	Subtraction Operator		Quotient Operator	
	Average Error	SD of Error	Average Error	SD of Error
0	15.702	21.779	13.197	17.943
1	7.105	9.949	3.413	4.186
2	8.354	12.571	2.325	2.877
3	8.904	15.577	5.839	12.173
4	5.244	8.259	1.734	2.579
5	11.529	15.825	7.873	10.915
6	7.099	11.301	4.611	6.146
7	12.204	18.730	7.015	14.232
8	8.706	12.348	2.603	3.555

Table 3.8



i) ii)

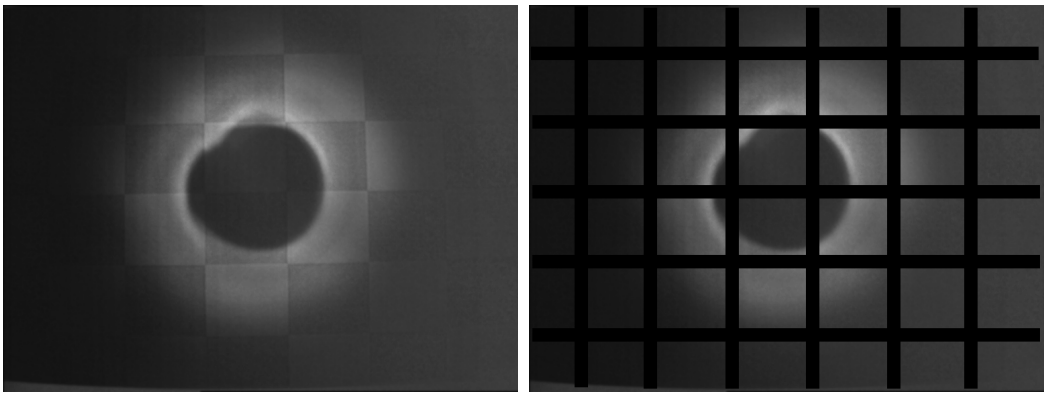
Fig 3.11 – Graphical comparisons of Average Error i) and Standard Deviation of Error ii) for Quotient and Subtraction operator results. Reading left to right, the first bar associated with each reflectance pattern represents results from the Subtraction operator, the second from the Quotient. Each level of filtering is represented by a different colour, the lighter shade of which being associated with results obtained using an averaged background.

Increasing the amount of smoothing up to this level, which is not excessive, produces a small but systematic reduction in reported error. In 83% of cases, the average error found was lower, where all input images had been filtered using a 7x7 Gaussian. With the exception of reflectance pattern 0, the largest mean error in the Quotient operator data recorded in Table 3.8 is 7.837 grey levels.

The most noteworthy outcome to be observed is that, in all cases, data processed using the Quotient Operator was less prone to error than that processed using the Subtraction Operator, sometimes by up to approximately three and a half times. Additionally, on average, the Quotient Operator produced results containing just under half as much error as the Subtraction Operator, with even less overall deviation. These results strongly suggest that the Quotient operator can produce descriptions of flashlight projections that are independent of the reflectance of the underlying surface.

There are some cases, however, in which both operators were found to be similarly prone to error. These, along with the best and most varied results, are examined by visual observation of the synthetic images they produce, the corresponding (real captured) ground truth images, and analysis of graphical representations of the difference images.

In order to allow unhindered visual analysis and comparison of the synthetic and ground truth images employed here it has been necessary to fit artificial black regions over any of the step edges encountered. This is because, for the human visual system, they have the unfortunate effect of creating an optical illusion (Fig. 3.12). If two planes of similar, or even identical, intensity taper even slightly in shade towards their boundaries, these can appear to form an artificially prominent edge, hence fooling the brain into believing both planes are very different in intensity (Aleksander 1987). This, in fact, may not be the case.



i) ii)

Fig 3.12 – Before, i) and after, ii) example of how applying gridlines counteracts the optical illusion of perceiving larger variations in surface intensity, either side of a step edge, than are actually present. [Example Image produced using background capture of reflectance pattern 8 (Fig. 3.8) in conjunction with the Quotient Operator]

The best result attained during experimentation (minimal variance and error), was obtained from images representing reflectance pattern 4. This featured a single vertical step edge marking a change from mid to high reflectance levels. As can be seen by visual comparison of the images in figure 3.13, with the Subtraction Operator's results (left), the estimation of intensity on the right hand side of the image is significantly too bright. Comparatively, the Quotient Operator's results (right) more closely match those of the ground truth (centre). The graphs in the same figure clearly show that, in fact, the results of the Subtraction Operator differ significantly on *both* sides of the image.

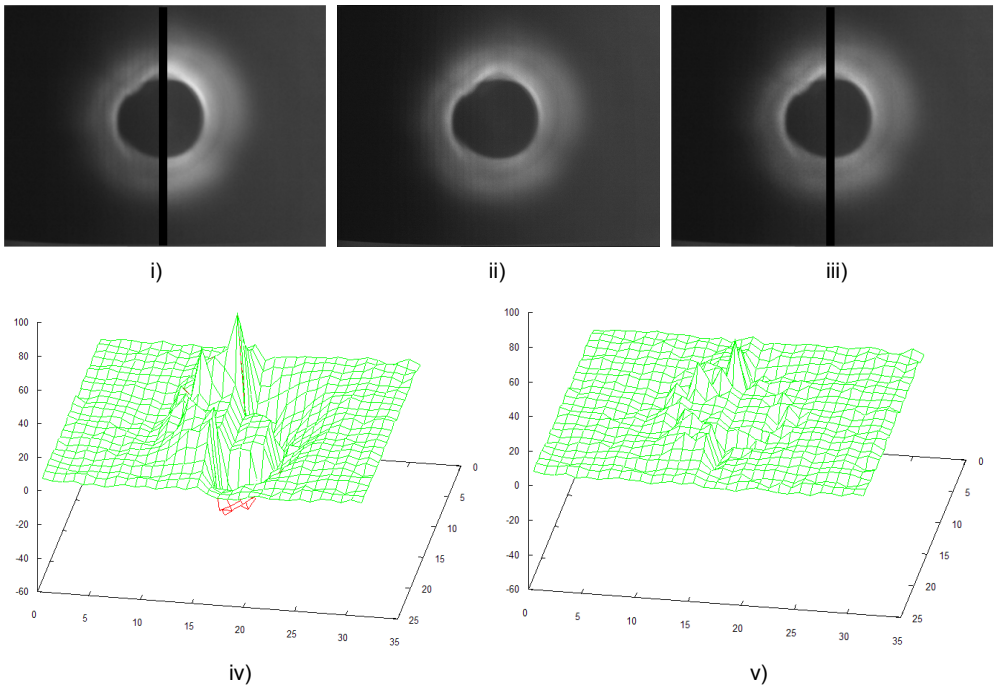


Figure 3.13 – Results gained from use of reflection pattern 4 (mid grey to white surface): i) Pseudo image produced using Subtraction Operator. ii) Actual captured image (ground truth). iii) Pseudo image produced using Quotient Operator. iv) Graphs depicting the calculated error of the results obtained using the Subtraction iv) and Quotient v) operator.

In this particular example the error produced by the Quotient Operator, although being slightly affected by the reflectance variations present in the image, is negligible. It in fact averaged at 1.73 across the entire image with a variance of only 2.56 intensity levels as compared to 5.24 and 8.26 for the Subtraction Operator. An interesting observation therefore is that, in the images produced from reflectance pattern zero (a uniform black surface), not only do we see the largest degree of error in the Subtraction Operator's results (overall) but also in those produced by its counterpart (Fig. 3.14).

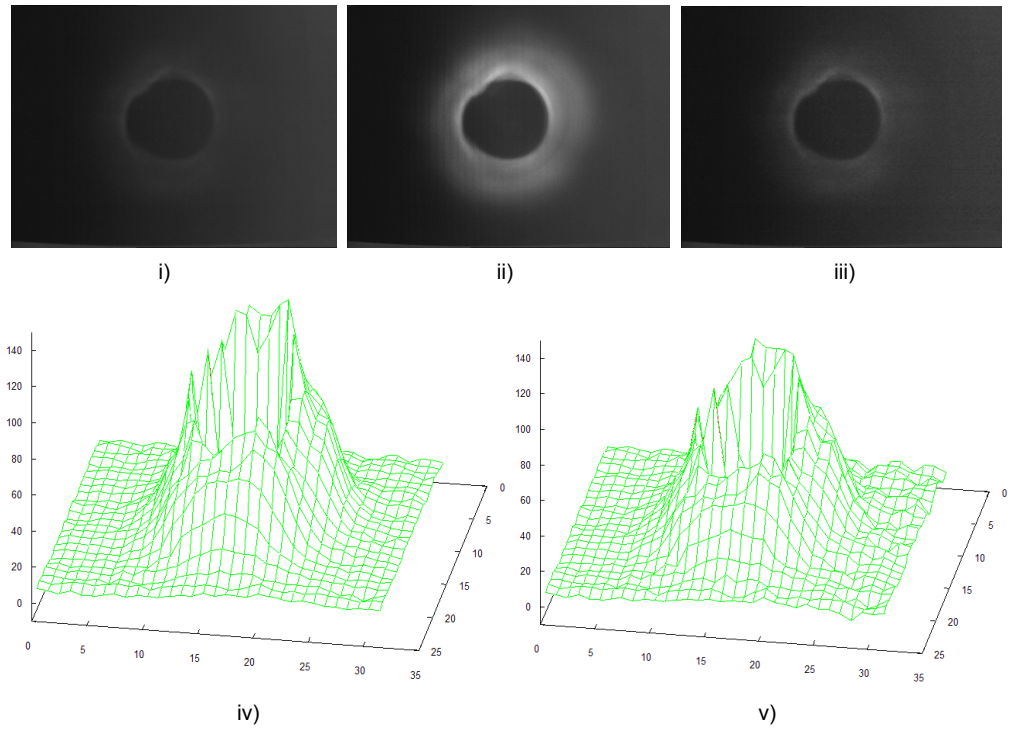


Figure 3.14 – Results gained from use of reflection map 0 (black surface): i) Pseudo image produced using Subtraction Operator. ii) Actual captured image (ground truth). Iii) Pseudo image produced using Quotient Operator, d) Graphs depicting the calculated error of the results obtained using the Subtraction iv) and Quotient v) operators.

On first consideration, such similarly poor results from the two different operators (mean error 15.7[S] and 13.2[Q], variance 21.78[S] and 17.94[Q]) seems inexplicable, since there were no reflectance variations present on the surface and these were expected to be the primary cause for instability. However, by observing the patterns formed from the error graphs, for each operator (Fig. 3.14 iv and v), it is clear to see that, in both, they closely resemble the pattern of light cast by the flashlight. The cause of the error is therefore likely to be poor digital representation of the flashlight, in the originally captured images. Given that the set, under scrutiny here, represents the darkest images used in our experiment; this indicates problems when working with particularly low intensity camera input. These problems are discussed further in Chapter 6.

This difference between Subtraction and Quotient operator results is most apparent in those generated from reflection pattern 8 (a grid of alternating surface variations from mid to most reflective). As is easily observed in figure 3.15, results gained from the Quotient Operator almost completely remove the effect of

surface reflectance variations in this case. Those generated from the Subtraction Operator however, still vary considerably from the correct response, as a result of them.

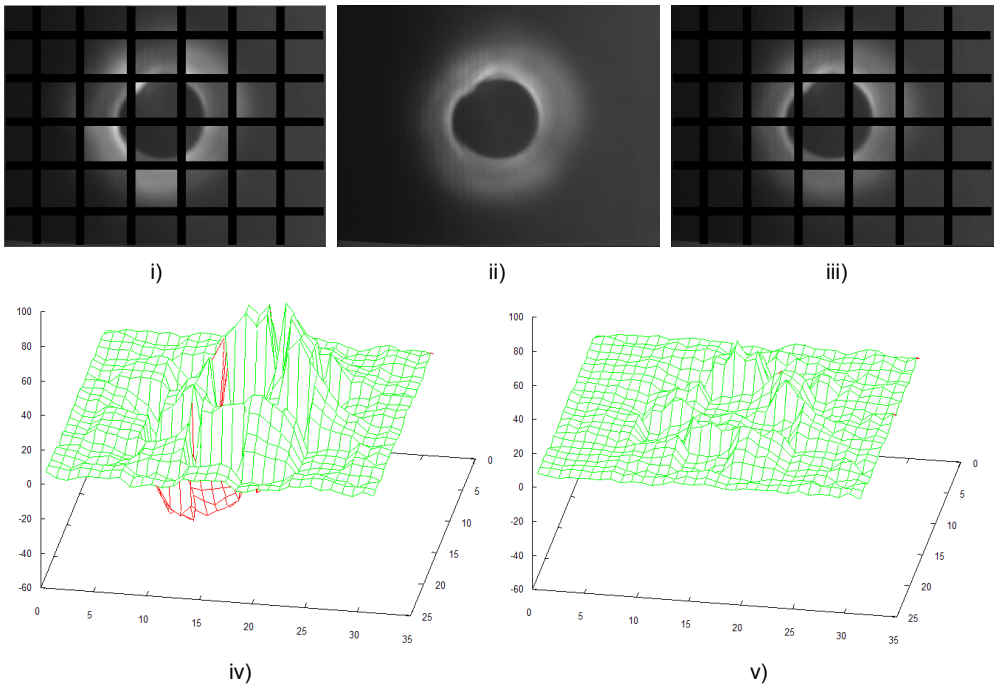
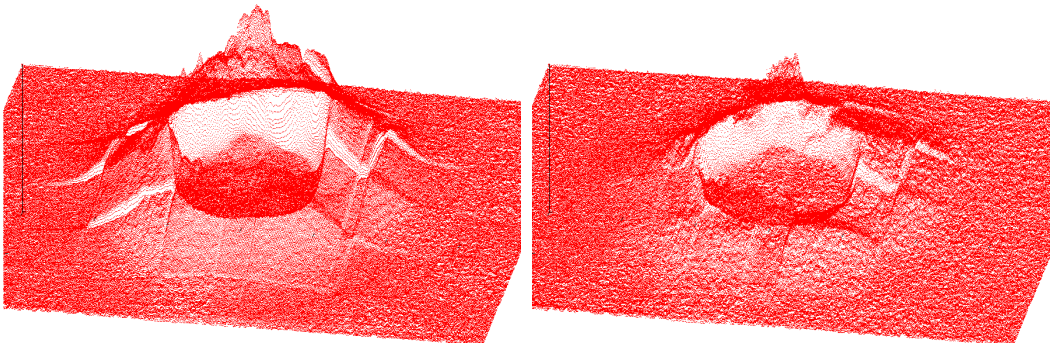


Figure 3.15 – Results gained from use of reflection map 8 (mid grey and white chequered): i) Pseudo image produced using Subtraction Operator. ii) Actual captured image (ground truth). iii) Pseudo image produced using Quotient Operator. Graphs depicting the calculated error of the results obtained using the Subtraction iv) and Quotient v) operators.

The results reported above focus on the relationship between mean errors (computed over the entire difference image) and the underlying reflectance patterns. It is interesting to ask whether errors vary systematically across the image, i.e. as a function of the illumination pattern projected by the flashlight. To this end mean errors were calculated at each image location from the nine error values recorded for each pixel in the difference images.



i) ii)

Fig 3.16 – Graphs (Scale 0-160) depicting the mean errors of difference images produced for: the Subtraction Operator i) and the Quotient Operator ii)

The values obtained (Fig. 3.16) are difficult to interpret; the errors measured are a function of the underlying reflectance values, which also vary across the image. Any conclusions drawn from this data may be specific to the set of reflectance patterns used. It is interesting to note, however, that the Quotient operator provides consistently lower errors across the region illuminated by the flashlight. This is more apparent in figure 3.17, which shows at which positions each operator's results exhibited less average error than the others'. Here, white pixels represent a lower Quotient Operator result and black, a lower Subtraction Operator result. It is clear to see therefore that the Quotient Operator provides the lowest degree of error (87.6% of the time) of the two operators. The Subtraction operator is only the more accurate of the two in regions where no additional (flashlight) illumination has been added. This is to be expected, since the Quotient Operator is known to be more sensitive to noise in low intensity images than its counterpart.

Figure 3.17 – Graphical representations of the average error of each pixel (over all results) to compare stability of the Quotient and Subtraction Operators. White pixels represent a lower error or deviation in results gained from the Quotient Operator and black from the Subtraction operator.

3.4 Conclusion

A variety of methods might be used to extract descriptions of flashlight projections from images. Early deployment of flashlights as interactive devices relied on standard techniques (thresholding and background subtraction). These were used in two contrasting situations. In the StoryTent, flashlights were viewed through a translucent screen. In the Sandpit and Nottingham Caves, flashlight projections were reflected from unaltered physical surfaces in the usual way.

Though Ullman's Source operator and methods developed for the identification of specularities appear relevant to the description of transmitted flashlight projections, examination has shown this not to be the case. Though suffering some limitations, no way of improving upon the threshold-based method used in the StoryTent has been found.

When reflected from a physical surface, the goal of a successful flashlight projection representation is to describe the projected illumination pattern independently of other factors contributing to the image intensity values used. Three approaches have been considered. Of these the Quotient operator is, in theory, the most promising. The values it produces are scaled only by ambient illumination, which can reasonably be expected to remain constant in many

situations and change only slowly in others. Experimental evaluation supports this conclusion. Though the Quotient operator is not always completely independent of surface reflectance (see figure 3.14 and Chapter 6) it has been shown to provide accurate and consistent representations of flashlight illumination patterns.

3.5 Summary

In this Chapter, the methods of extracting descriptions of flashlight projections from images captured in the circumstances described in Chapter 2 have been examined. The threshold-based approach used in the StoryTent appears to represent the best available method when flashlight projections are transmitted to the camera through a screen, rather than reflected from a physical surface. Transmitted flashlight projections will not be considered further. When flashlight projections are reflected from a physical surface, the proposed Quotient method has been found to provide accurate representations that are independent of the reflectance of the target surface. This method will be used throughout the remainder of the work reported here. In the next chapter, the deployment of a new flashlight interface (based on the style of use in the Caves), in five experimental installations, is discussed in full. Emphasis is placed on highlighting each location's technical requirements and challenges.

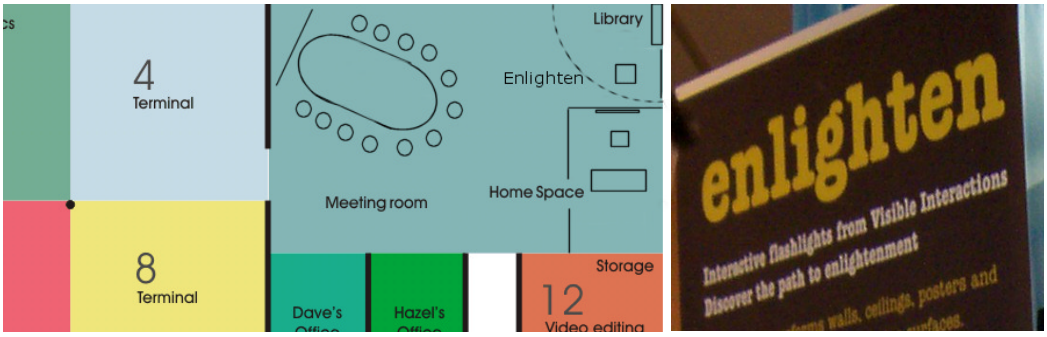
Chapter 4

Experimental Environments

The work described in Chapter 2 determined that the use of flashlights as interaction devices in the tent, sandpit and cave scenarios had merit but would benefit greatly from the additional functionality of being able to recognise different flashlights. This would provide richer, more varied, content for these user experiences. Future extensions based on uses of recognition (for example, to control sound playback) or exploitation of other features like quantifications of a flashlight's degree of rotation (which can potentially be extracted from a better representation of its beam incident on a surface) would also be possible. In Chapter 3 a technique was conceived, developed and tested which allows us to extract a representation that is, although not an exact measurement of the illumination from a flashlight, accurate enough (under most circumstances) to depict its distinguishing features. This can be done with enough clarity that the individual light becomes identifiable through pattern recognition and other features can be extracted. This method forms the core of the proposed system. Any interactive device based on such a method is, however, likely to be required to operate in a wide variety of situations. In this chapter we examine in detail five experimental environments for which extensions or modifications of the system originally installed in the Nottingham Caves, have been developed. Each makes use of a form of flashlight recognition that exploits the extraction/representation technique described in Chapter 3 but also raises a unique set of issues which are noted at the end of each section. In the following chapter we shall examine in detail the components that make up the flashlight system, used in these installations. Reference is made to how each of the issues described here have been addressed.

4.1 Laboratory-based Demonstration and Development Rig

The Mixed Reality Lab (MRL) in Nottingham University's School of Computer Science incorporates a large working area used for research projects spanning a wide variety of disciplines. It can be divided into a number of bays; smaller working areas that are used for experiments or technology development. Together with those areas that experience a more dynamic use, several bays feature, either technological showcases demonstrating examples of work produced by or in association with members of the MRL and its connected research groups (Craven 2001), or long running experimental installations (Schnädelbach 2006). One such bay (Fig. 4.1) features a semi-permanent installation of a flashlight demonstration whose purpose is threefold, serving not only as a rig for development, testing and analysis but also as a lab technology demonstration and for marketing Enlighten, the commercial version of the flashlight system, to potential customers.



i) ii)

Fig 4.1 – i) MRL floor plan depicting location and layout of "Enlighten" demonstration. A wall display is augmented as is typical in schools. The dashed line indicates a curtained area. ii) Marketing and Exhibit Information.

The installation in the MRL is based around the augmentation of the type of displays often assembled by primary school children to represent the knowledge they have learned on a particular topic or the work completed over a certain time period. In this particular case, the theme of the display (as pictured in figure 4.2) is 'The planets of our Solar System' which originally featured painted cardboard cut outs of each planet together with the sun and earth's moon and was created by Nottingham primary school children (Green 2004). To augment the display, content tailored separately to both adults and children is employed and, to this end,

the system is configured by default to recognise two separate flashlights. The first of these flashlights (often smaller and lighter to make it more suitable for children), when shone on the various target planets, triggers narrated content recorded by the children who made the display. For example: *"Hello hello. This is Jupiter; watch out for the red spot!"* The second, so called, 'adult flashlight' (often larger and heavier) instead plays associated extracts from Holst's Planet Suite. The choice of separating voice and music as associations to individual flashlights has the further advantage that both flashlights can be used simultaneously, either by two users or a single user (with one in each hand as in figure 4.2i), without the two groups of audio clips conflicting with one another (see included video).

i) ii)

Fig 4.2 – Original planets display created by school students i) as compared to the version used for commercial demonstration purposes ii). In addition to the use of professionally designed art work, the latter display is both portable and flexible as (regards how it can be arranged) which is used to demonstrate the ease of configuration of the software. It also minimises the high background contrasts found in the original.

The bay in which the MRL flashlight demonstration is installed is a 3x3 meter region of lab space that can be curtained off by a circular rail from corner to corner. Along one wall is a row of large windows obscured by blinds and along the other, the planets display consisting of a 1.5 x 1 meter region of light brown wooden panelling featuring a number of two dimensional planet representations to the right of a book case. The area is monitored by a CCD camera fixed to the ceiling 1.8 meters away from the wall above a table where the flashlights are placed. This provides an implied (but not enforced) barrier between the user and

the display which, for the most part, prevents the camera's field of view being occluded by a potential user.

i) ii)

Fig 4.3 – i) Illumination is partially controlled by strip lights, blinds and curtains; however variations in natural light still have significant effect. ii) A variety of flashlights can be tested and demonstrated with this display including head torches and a modified version of the spotlights used at MAGNA (see section 4.5)

Lighting is provided from both fluorescent strip lights and natural light entering through the windows as demonstrated in Figure 4.3. Due to the presence of blinds this can largely be controlled and kept relatively constant, however gradual changes in lighting levels, over the course of a day, and those due to extreme changes in weather (e.g. clear sky to dark cloud cover) do have a significant effect. Illumination, across the area monitored by the camera, is however approximately constant as required for the successful operation of the Quotient Operator detailed in Chapter 3. This is achieved by use of the bay's curtains in screening off light and shadow that may be cast on the wall from neighbouring, non-controlled, light sources. The current version of the display features planet representations that minimise strong contrasts with the background (Fig 4.2).

As regards the MRL installation, there were no specific deployment requirements due to both the flexible nature of the demonstration and the fact that it is not required to be operational full time. The program interface for running and

configuring the system is only intended for use by developers and experienced users and there is only need for processing of one camera's input at a time. Additionally, unlike other installations, there is no need to hide or disguise the equipment/technology (although the means exist to do so) as it is often beneficial for visitors to be able to see how the demonstration works and get a behind the scenes look at the processing involved in driving the system.

The identified technical issues relevant to this installation are:

- General ambient illumination level changes over time
- Possible (but unlikely) occlusion of the camera monitored area by users

4.2 Newark Agricultural Show

The Newark and Nottinghamshire County Show (see website) is an annual event organised by The Newark and Nottinghamshire Agricultural Society at the Newark Showground in the East Midlands which, whilst featuring a number of public spectacles (monster trucks, motorcycle display teams etc), mounted games, competitions and livestock displays serves also as a large trade event for businesses exhibitors and draws over 55,000 visitors annually over its two day weekend duration. The University of Nottingham is a regular attendee at the show whose presence there takes the form of a large semi-permanent rigid multi-segmented marquee featuring a number of exhibitions and demonstrations of work being undertaken by its various researchers and schools. One such demonstration was of the flashlight technology as a representation of some of the work being undertaken by the School of Computer Science's Mixed Reality Laboratory.

The Newark Show demonstration was based around a children's story telling and drama activity/workshop scheduled and run as a fixed number of sessions each day by Rachel Fenely, a community development officer employed by Nottingham City Council to work with, and organise events for children in the

local area. These events have a variety of purposes, but a major aim is to encourage the growth of the children's creativity and imagination. The theme of each workshop was a loosely planned adventure-based storyline called "The Journey into Space" which featured the participating children imagining and acting out the activity of going on a mission to the stars. Along the way, an interactive flashlight display, similar in nature to the one installed in the MRL (see section 4.1), featuring use of 2 flashlights to illuminate planets and other targets, was used to help guide the experience. A significant difference however, to the static demonstration in the lab, was that as part of the storyline the children were participating in, they themselves recorded the sounds which were then used in the configuration of the system. Later on, when the children came to using the flashlights, the sounds they heard were a mixture of pre-recorded effects and those that they themselves recorded.

During the course of the 2 day demonstration some 19 sessions were run with groups of 2-6 children largely in the 3-7 age range. The story was played out by making use of two physical spaces, one containing the flashlight technology and the other a rug for the children to sit on. A significant aim of the drama experience was for events to be driven largely by the children's imagination with Rachel's role being primarily to improvise and work the story around the ideas and events that the children themselves envisaged happening. Despite this deliberately varied experience, each session loosely followed the following preconceived order of events:

1. The first stage of the adventure began in the *flashlight area* (top half of figure 4.4) where the children, or 'space cadets' as their role dictated they be referred to as, saw and discussed the planets, alien and spaceship on the "talking wall" (as it was referred to during the experience) and were asked questions by Rachel such as how they thought the aliens would communicate, which planet (on the wall) they might live on and what noises their own space ship would make. Once these had been decided on and rehearsed the children gathered round a microphone to record them. Typically three sounds were recorded per session which included a countdown to take off, alien sound effects and an emergency warning from the 'captain' (usually one of the children) ordering all cadets to return to the ship.

2. While a computer operator went about saving and attaching the newly recorded sounds to targets using the flashlight systems' interface, the cadets retired to the rug on the other side of the wall (*preparation area,* bottom half of figure 4.4). Here they completed their 'training' which involved, among other things, learning an alien dance to use when meeting extra terrestrials and practicing walking in low gravity environments.

3. To begin the mission itself, the group returned to the *flashlight area* and boarded their spaceship by sitting in a circle. 'Blast off' was initiated by use of a flashlight or flashlights on first, their recently recorded 'countdown target' (earth) followed by another target (or use of a different flashlight on the same target) that triggered a stock recording of a rocket ignition sequence. The cadets then travelled through space looking for their destination by taking it in turns to shine a flashlight out of the ship window toward different planets.

4. Most of the planets they 'discovered' would trigger audio recordings taken from the MRL based demonstration discussed in section 6.1 but the crew were really in search of a special 'Alien Planet' which, when illuminated would trigger the children's own alien sound effect, identifying it as the target for their destination.

5. Once found, the spaceship landed and the group would then explore this planet looking for items such as healing rocks or gold to take home. Along the way it was common for them to encounter an alien (usually a spectating parent) whereupon the cadets attempted to make contact by performing the alien dance they learnt as part of their training. The 'aliens' invariably turned out to be hostile where upon, further use of a flashlight triggered the recorded message from the captain for the group to "return to the ship!"

6. The cadets would then endure a rapid 'Blast off' (as in stage 3), make the journey back, and return home to be greeted by much applause and appreciation from a grateful world.

The area in which the activity took place was a large hexagonal "pod" segment of the aforementioned marquee whose construction consisted of three such pods, the remaining two being separate, and used to house unrelated demonstrations. The pod itself (Fig. 4.4) was constructed of five 5x2 metre wall sections (the sixth side being open to connect the pod to the rest of the marquee) two covered by high quality sheet graphics depicting planet and spaceship themed images in the "pop art" style and the remaining three covered in plain black material save for a

number of speech bubble style boards that detailed information about the exhibit to visitors. To cover this area, a roof was drawn to a point approximately 5 metres above the centre of the hexagon which was made of a light, neutral covered canvas. The division of the pod into the two areas required for the experience, was achieved using a sixth wall, again featuring planetary art work, which was positioned parallel to the entrance, in the centre of the pod, with room around each side for walkways.

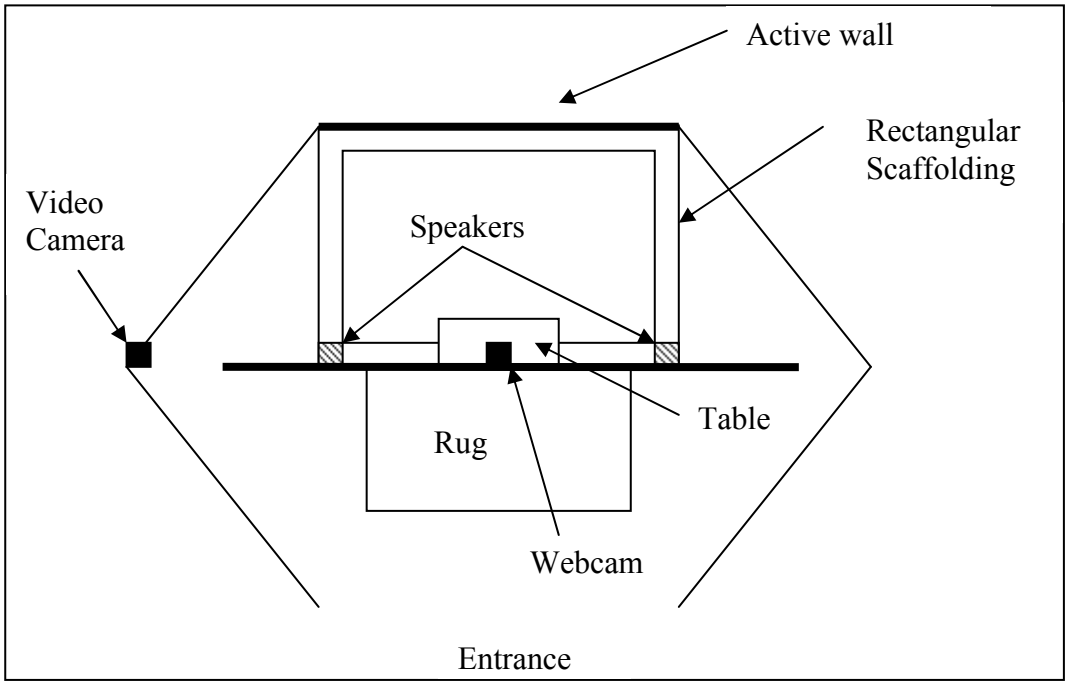


Fig 4.4 – The two areas used for the story telling/drama experience in the "Journey into Space" pod. The upper half shows the positioning of the speakers, camera, table and active wall in the *flashlight area*. The bottom half shows the, technologically devoid, *preparation area* where the participants received their 'training'.

In order to deploy computer equipment and hence make the wall segment opposite to the enclosure's opening 'active', a rectangular girder/scaffolding (as is typically used in lighting rigs) made of light weight aluminium tubing was installed on the opposite side of the centre wall to the rug, which extended all the way to, across the top of and down each side of this active wall. To monitor the wall, a standard domestic web cam was mounted on the scaffold opposite the active area and USB extension cord was fed through it, down to a laptop computer that was located on a trestle table directly beneath. Unlike the MRL installation, it was necessary here to attempt to hide the equipment and, to this end, a 3D representation of a Jupiter-

like planet was constructed from pressure board, beneath which it was possible to conceal the computer, cabling and amplifier which was used to power the floor standing speakers placed on either side. The targets themselves (which subsequently became part of the MRL installation described in section 4.1) consisted of large (50-75cm diameter) cut-out planets of a similar style to the aforementioned wall graphics constructed from the same material as was used in the table cover. In addition to these planets, a number of cartoon images (aliens, spacemen, treasure etc), provided by Rachel herself, were also utilised and these were affixed to the active wall and its surroundings, using Velcro, as depicted in the photograph in Figure 4.5.

Fig 4.5 – Rachel and “Cadets” explore space, using “magic torches” from within their imaginary Rocket Ship.

Although the space did not feature any window fittings and the like, the aforementioned canvas roof was not opaque, and its translucency provided much of the illumination available in the space. Not only were the light levels in the pod much brighter than in previous installations (such as the Caves and MRL), during periods of clear sunshine, but they were also highly changeable, a sudden

occurrence of cloud cover having the potential to completely alter the visibility of a torch in use at the time. Further issues were also introduced when considering the suitability of the web cam that was used as such equipment provides only limited control of camera parameters such as frame rate and exposure which hindered configuration under these conditions. Additionally camera positioning in the installation was not ideal as the webcam could not be placed at the optimal distance (so as to maximise relative resolution of flashlight beam representations) from the monitored wall. This was due to a lack of a central scaffold on which to mount it. Finally, there was an undesirably high contrast between the relative brightness of targets and the dark colouring of the material to which they were fixed.

Like the MRL installation, the software for the "Journey into Space" demonstration was only to be used by an experienced operator and therefore no customisation or changes to aid usability were required. One side effect to this however, was an unavoidable break in the flow of the experience due it being hard for Rachel and the participating children to remain in character while recording audio (as per stage 1 described above). A considered solution to this problem was the creation of a, so called, "interplanetary radio" that might take the form of a spaceship like control console featuring a button the children could press in order to start recording. Although such an interface would have greatly helped with the aim of hiding the technology, no such customisation of the software was attempted as this would have required, in addition to the ability to record audio clips, the need to also automatically level and remove silence from them and make association with each of the relevant targets. Due to the non-permanent nature of the installation, implementing this level of interface complexity was deemed inappropriate however the consequential requirement for this processing to be completed manually, shaped events leading to the space training phase of the story becoming a vital part of the overall experience.

The identified technical issues relevant to this installation were:

- A significantly bright environment requiring the use of large and powerful flashlights in order for them to be visible to users and spectators
- Highly dynamic and rapid changes in illumination levels
- Non software controllable exposure and frame rate settings on camera equipment (commonly leading to over or under exposure of imagery)
- High relative contrasts between targets and background material
- Low resolution images (due to non-optimal camera positioning)
- Occlusion caused by members of the public and participants obscuring the camera's field of view

4.3 The Flint Kiln, Etruria Industrial Museum

The Etruria Industrial museum in Stoke on Trent is a small, family oriented museum open afternoons for three seasons of the year that is home to the *Jesse Shirley Etruscan Bone and Flint Mill*, the last surviving steam powered potter's mill, of it's kind, to be found anywhere in the UK. The main focus of the museum is of course the mill itself (Fig. 4.6). Built in 1857 to grind raw materials for the pottery and agricultural industries on the junction of the Trent & Mersey and Caldon canals, the mill ran in the traditional fashion for 115 years till it was eventually shut down in 1972 and then scheduled to be an ancient monument in 1975. Despite its age, the historic machinery is still fully intact and is regularly demonstrated, running in full steam on a number of weekends throughout the year. In addition to the mill, the museum also comprises a community visitor centre featuring an interactive hands-on exhibition that tells the story of Etruria, the pottery industry in the eighteen hundreds and the mill itself. Tours are scheduled daily whereupon visitors can enter the mill, see how it worked and learn about the hardship and dangers of working there during this period in history.

Fig 4.6 – Etruria Industrial Museum (Aerial Photography – Microsoft Live Local)

A prominent feature of the tour and the museum itself is a large (building sized) kiln, located between the visitor centre and the mill, which was used for the calcining of raw materials such as flint. It was for this particular exhibit that the proposal of augmentation using interactive flashlights was put forward.

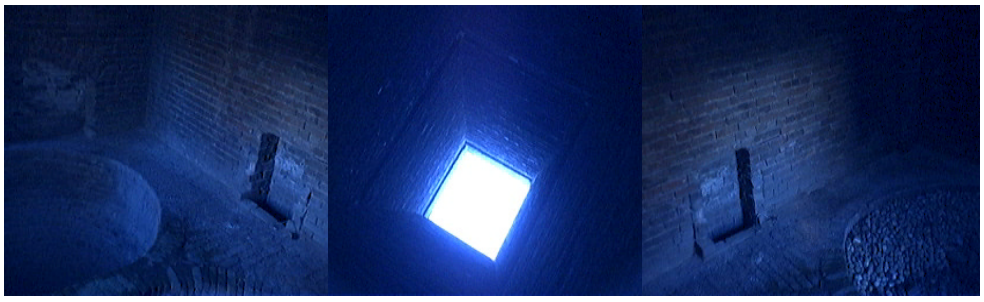


Fig 4.7 – The dark interior of the kiln as viewed by the two installed cameras (left and right images) Internal illumination is provided by natural light from both the chimney (centre) and the door way (not shown).

Prior to installation, the kiln itself was experienced as a single roomed building into which the public could enter and observe its interior and features while standing in a small area located just inside the doorway behind a barrier to prevent any risk of falling into one of the exposed pits. The kiln itself offered no explanation as to its purpose or history at this time save for a small plaque attached to the outside or the information provided by an experienced guide should it be visited as part of a tour. To augment the kiln, the proposal was to install flashlights affixed to the barrier which the public could use to aim at targets again, in order to trigger play back of audio clips which were designed to provide explanations as to the history and purpose of its various parts.

Flashlights are particularly suited to this environment since, not only is the interior relatively dark, it being lit only by natural light from its chimney (see Fig. 4.7) and doorway, but also because the kiln is listed under the National Trust (see website) as a protected building. This, although safeguarding it against potential damage or alteration, means that no tags or placards may be fixed to the walls leaving non-invasive augmentation techniques as the only practical solution for providing information to the public.

The layout of the kiln, as shown in Figure 4.8 is a rectangular room with the angled barrier approximately 1.2meters high that creates a triangular area of about 1.5m² inside the door in which the public can stand. On the other side of the barrier, two large egg-shaped circular pits are built into the floor, one filled in with exposed layers of flint stone then coal repeating and the other empty to reveal a shaft at its base. In the lower centre of the opposite wall is an opening which was used as an exhaust flue (Figure 4.7 left and right images). In the far corners, a propped up wheel barrow containing cattle bone calcite, with a pile of bloody bones placed next to it, and a mine cart full of flint pebbles are located respectively (Fig. 4.8 top). These, although not originally features of the kiln during its operational life span, were added to provide context and visual clues as to aspects of its historical use. The chosen targets were therefore: the calcite, the mine cart, the flint/coal layering, the wall of the empty pit and its base shaft

(diagram, Figure 4.8). In addition to these obvious eye catchers, a further two targets were positioned on featureless sections of the far wall. The reason behind this choice, although partly to avoid use of too many cameras (it being already necessary to utilize two, just to cover the area and angles required to view the entire kiln and bottom of the pit) was mainly due to experience gained during the Nottingham Caves demonstration (Ghali 2003). At the time, use of so called 'hidden targets' had prompted strong positive visitor feedback and so it was deemed beneficial to the installation, to employ this strategy once again.

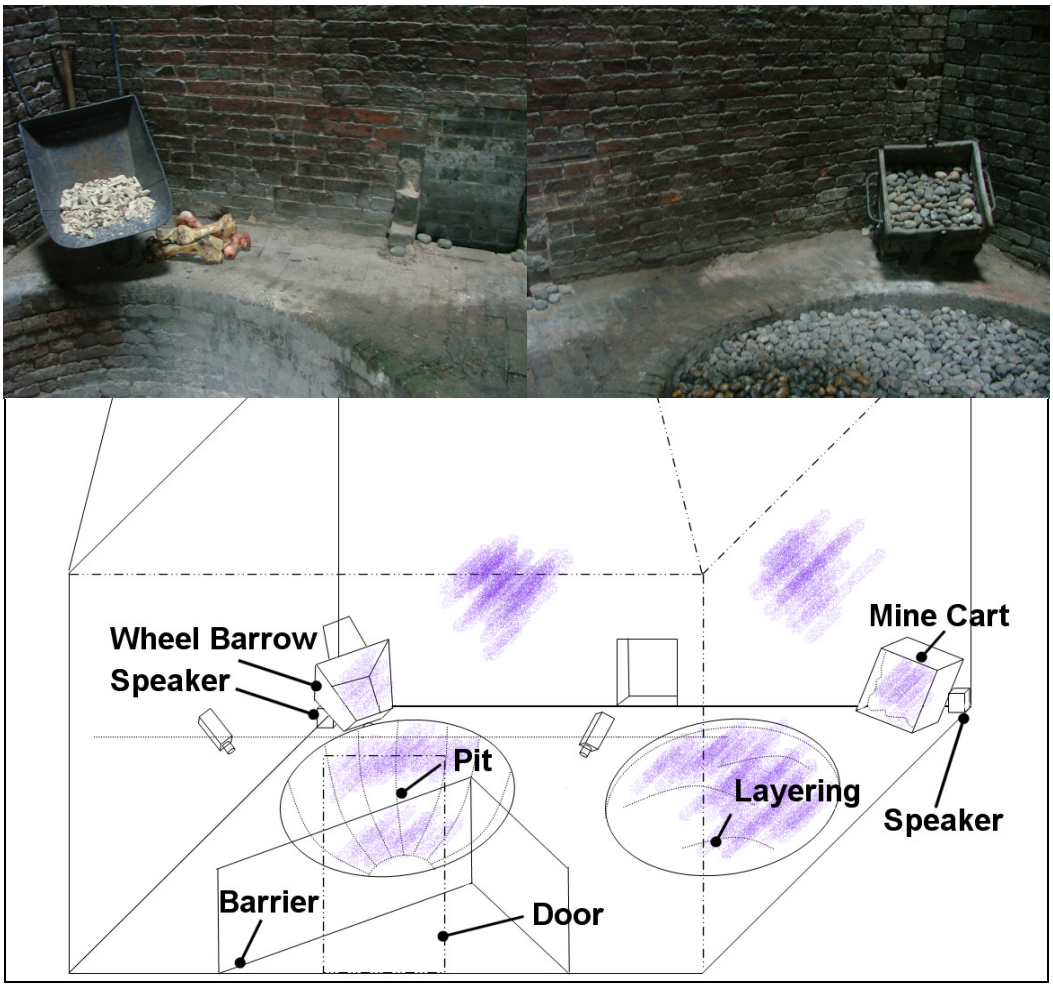


Fig 4.8 – Mock 3d layout of the inside of the kiln. Visitors stand behind the labelled barrier and direct flashlights towards any of the target areas marked in purple. Above them, cameras which monitor the area are fixed to a horizontal pipeline and sounds are played back through hidden speakers. Top left: The calcite and shaft targets. Top Right: the mine cart and layering targets. Natural dust and stones are used to disguise conspicuous cables.

The lighting in the kiln, it being exposed to the sky, has much in common with the marquee used at the Newark Show (see section 4.2). Variations in the general level of illumination can be vast and occur over very short periods of time in

accordance with weather and cloud covering during the day. In addition, a significant portion of the light incident on the monitored area is provided from the doorway which, when obscured by visitors, can not only cause strong changes in the overall level of illumination in the kiln but also produce more specific lighting features such as soft shadows on the targets and opposite wall. The targets themselves, many being three dimensional in nature, can also be self occluding and consequently may cast strong umbrae on themselves or the space behind them when illuminated by the flashlights. Additionally, unlike in previous installations where the monitored area is flat and approximately perpendicular to the expected point of origin where the flashlights are used from and the camera positioned, most targets in the kiln are either illuminated by flashlights, or viewed by their allocated camera at an angle that is comparatively extreme. Due to the high positioning of the cameras (above, behind and to either side of the publicly accessible area) and the installed barrier in front of the doorway, occlusion of either camera's field of view due to a visitor's actions would require an extigent effort to achieve and is therefore unlikely to occur.

Deployment of the Enlighten system in the kiln was the most difficult of the installations discussed so far, as there were a number of restrictions, constraints and environmental factors to consider. These stemmed mainly from the aforementioned protected status of the kiln (meaning the installation cannot, in any way, be permanent) and it being, effectively, an outdoor location. Conditions in Etruria's kiln are cold, dusty and damp, largely attributable to the proximity of the canal (mere meters away) and the kiln's roof (chimney) being open to the sky leaving the interior largely exposed to the elements (despite narration in the audio clips used and the museum curator claiming the tapered walls prevented this). This environment, in addition to making working conditions during testing and configuration stages difficult (snow for example was a factor, see figure 4.9iii), also required the weather proofing and protection of all equipment to be located within the kiln interior. The equipment itself had to be suitably rugged enough to withstand such extreme conditions and continue to function all year round. The cameras used are therefore stock CCTV devices rather than web cams (conditions,

cable length limitations and low illumination levels making these impractical) which are placed in custom made housings and attached using adjustable camera mounts clamped to a pipe running the length of the back wall above the door (Fig. 9i). Custom housings were needed, again, due the limitations regarding permanent fixtures and the mounting space available being too restrictive for even the smallest commercially available solution (Fig. 4.9ii). Other ruggedisation requirements for the kiln installation include the use of thick power/signal coaxial cable runs for the cameras, to avoid possible damage by rodents, and water proof speakers.

Fig 4.9 – i) Cameras fixed to an original pipeline within custom made housings designed both to protect and camouflage them. ii) The equivalent smallest available commercial solution is too large, heavy and conspicuous to be suitable for use at Etruria. iii) Cold and snowy working conditions within the kiln

The computer equipment employed to run the software is not located in the kiln itself, partly due to power and space requirements but primarily so that the exhibit appears completely unaltered and devoid of technological enhancement. For this reason, to aid in hiding the technology, the speakers were placed out of sight behind the corner targets (a wheel barrow and mine cart) and every effort was made to conceal cables using natural dust or stones from the surroundings (see Fig. 4.8) and by working them into gaps between brickwork in the areas that were deemed least conspicuous to visitors. The computer is installed on a narrow shelf concealed behind a fake doorway along with the controls for other nearby exhibits. This is located in an area of the mill known as the Engine Room which is directly beneath the installation area and requires a cable run of approximately 30 meters that exits the kiln via the aforementioned exhaust flue and is hidden by machinery until reaching this control point (Fig. 4.10). To fit the cupboard, a particularly low profile PC has been utilised along with a touch screen flat panel TFT monitor

and suitable user interface in order to compensate for the lack of space available for use of a keyboard or mouse respectively (Fig. 4.11). The physical distance between the computer itself and the area in which the flashlights are actually used also presented a challenge during the initial installation of the software. Due to a need to observe the cameras' output (for example to focus them and adjust aperture) at the configuration stage, both post and prior to processing, it was necessary to utilise a two person team communicating via CB radio in order to achieve optimal performance in an efficient manner.

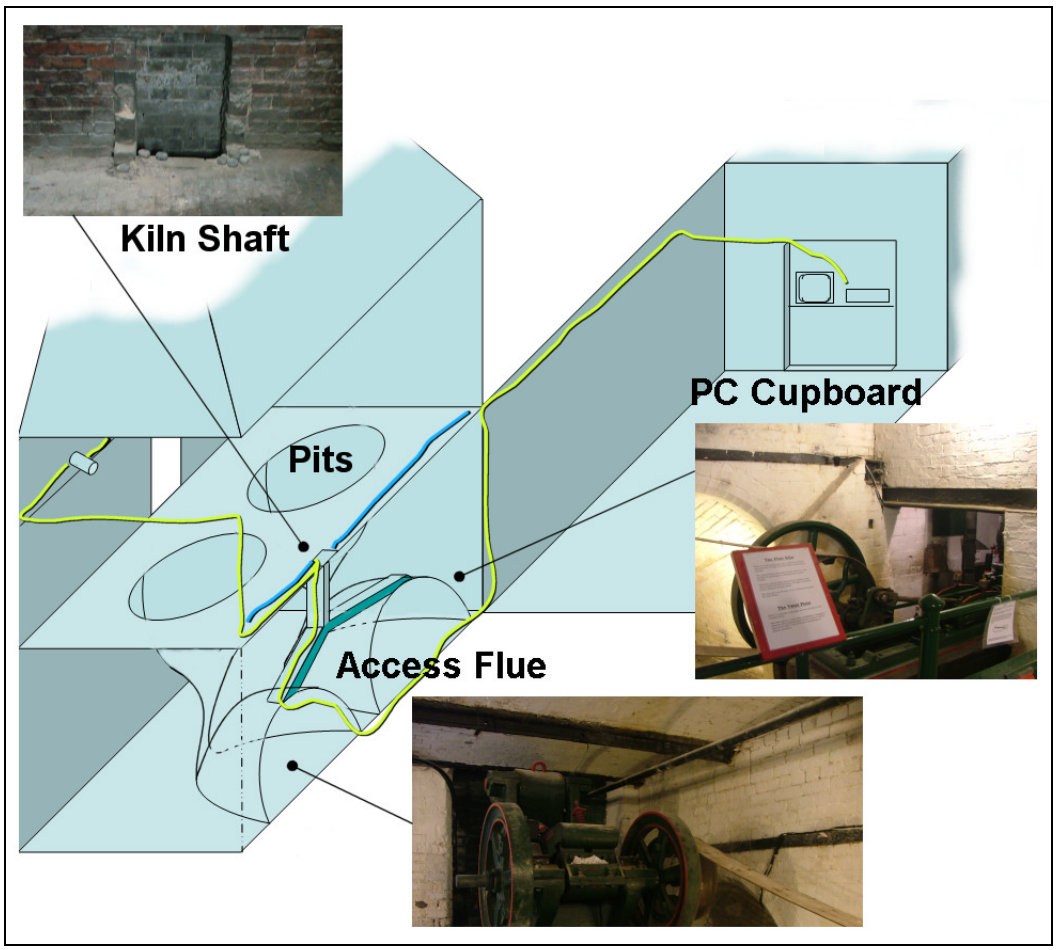


Fig 4.10 – Access, space and historical preservation restrictions required computer equipment to be installed in another part of the mill. To avoid visibility, the camera (yellow) and speaker (blue) cable runs followed a long complex route (between brickwork, down an exhaust flue and behind pipes) to a storage cupboard.

As regards user interface, Etruria, being the first museum exhibit to make use of Enlighten, had very different requirements from those of the previous installations. At its most basic level the system had to be capable of full automation, meaning the computer boots, runs the Enlighten software and shuts down on a schedule

without any need for further configuration or adjustment stages by a museum curator or engineer. Additional constraints governed by Etruria are that, in the event of failure, the system can be reconfigured via a very simple step by step interface that is suitable for non-computer literate users (see Appendix A).

Fig 4.11 – The computer and touch screen interface, normally hidden from view, as shown on right.

The identified technical issues relevant to this installation are:

- Highly dynamic and rapid changes in illumination levels
- Non spatially uniform variations in illumination
- Avoidance of over or under exposure of camera imagery attributable to lighting changes.
- Three dimensional, self occluding targets against non planar backgrounds
- Variations in surface properties over time due to precipitation (e.g. rain increasing reflection and snow accumulating)

4.4 Intech Science Centre

Intech Science Centre is the south of England's premiere hands-on interactive science and technology centre. Located in a 3500 square foot, award winning building (Fig. 4.12) housing over 100 interactive exhibits designed to demonstrate the science and technology of the world around us in an engaging and exciting way, the centre is a popular and regularly visited location for families, schools and

other local education establishments. Currently, the millennium commission funded enterprise has over 80,000 visitors per year with over a quarter of these accountable to school field trips and its popularity is rapidly increasing by more than ten percent annually. This high percentage of school visitors is due largely to partnerships with both a number of local schools and also the local education authorities. As a result, many of Intech's interactive exhibitions and related workshops have become integral parts of curricular and extra curricular learning within the region.

Fig 4.12 – The futuristic Intech building and site in Winchester

Together with the learning benefits of interactive exhibits (Ramey-Gassert and Walberg 1994) that serve to illustrate and strengthen established areas of the curriculum, Intech also aids in the public outreach efforts associated with scientific projects and developments that can be tied to but also go beyond the current national curriculum. Examples include optical digital sky surveys like the Sloan Digital Sky Survey (SDSS) and the United Kingdom Infrared Deep Sky Survey (UKIDSS) which are generating vast quantities of information detailing the intensities and positions of hundreds of millions of galaxies, stars and quasars which help to further our understanding of the universe. The surveys provide a constant source of emerging discoveries which are shared by scientists world wide and so, with an aim to help promote and raise awareness of this work, Intech proposed to use material from these surveys in order to create a learning experience that is not only exciting and inspirational, but also meets and goes beyond the National Curriculum elements relevant to this field of science and exposes pupils and the general public to our developing knowledge of the universe.

The proposal put forward was to make use of interactive flashlights to create an interactive exhibit aimed primarily at 11-16 year olds (approximately 5,700 were expected to visit Intech within the 1st year of operation) but which is also flexible enough to appeal to all visitors while still providing school pupils with extra curricular information beyond what is normally found in key stages 3 and 4. This would allow visitors and pupils to carry out their own exploration of the subject using different flashlights to discover a range of facts and theories at varying levels of complexity. The constructed exhibit, entitled "Explore the Universe – Learning through Interaction" features educational material specifically selected to cover National Curriculum Science Elements "The Solar System and Beyond" (QCA7L), "Gravity and Space" (QCA9J) and sections of the sub module, "The Universe and how it continues to change" from within GCSE Science and Astronomy. Due to the changing nature of the curriculum and of course to allow for the inclusion of new discoveries and data from the sky surveys as they occur, it is the intention of the Intech curators to utilise the flexibility of Enlighten in order to easily update the exhibit content in the future. This will ensure that it remains both educationally relevant and on the cutting edge of current scientific knowledge for the duration of its lifetime.

Explore the Universe consists of a large, semi-enclosed oblong space approximately 10 metres wide, 2.5 metres high and 2 metres deep within which can be found four stand alone installations of the Enlighten interface (Fig. 4.13). Enclosing the exhibit makes the ambient illumination approximately constant, and removes the potential problems caused by the large number of large windows in the Intech building. Internally each installation is arranged so that its flashlights are intended for use over only their associated parallel section of back wall and it is expected that visitors typically enter at one of the two open ends of the exhibit then gradually work their way through trying each installation in turn. Although the content for the four wall segments are effectively standalone, movement between them, in a linear fashion, starting from either end of the exhibit also represents a natural and logical flow of information.

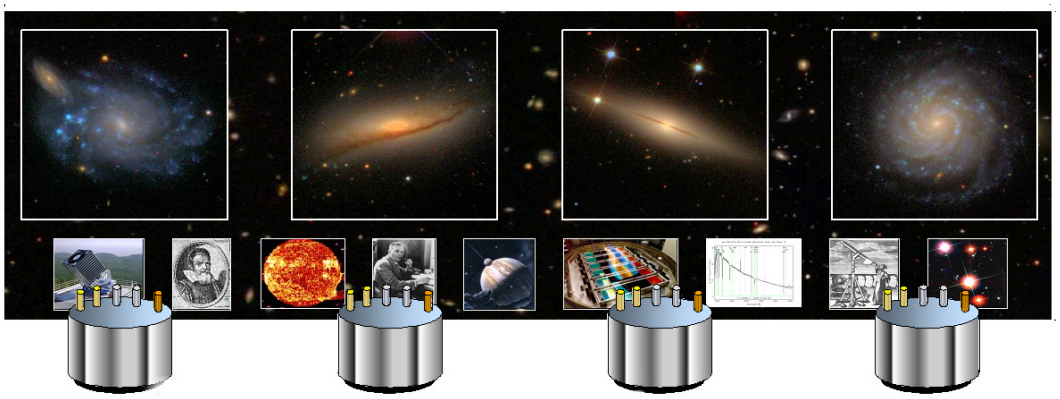


Fig 4.13 – Concept diagram for original Intech design

Primarily, the draw for the exhibit is intended to be the public's inquisitiveness as to what lies inside its seemingly dark interior (Fig. 4.14). The walls (both internal and external) feature pseudo colour imagery of the night sky obtained from scans by the SDSS and, additionally, the outside surface displays a number of traditional information boards containing text, diagrams and images that explain: The Sloan Digital Sky Survey and the science behind it, the SDSS telescope and the complexities of its specialised CCD array, how telescopes can be used to see into the past, the measurements of time and distance and, finally, details regarding the Apache Point Observatory where the survey is located.

Fig 4.14 – The semi enclosed *Explore the Universe* exhibit acts both as a draw and serves as an introduction to the Sloan Digital Sky survey. Visitors are compelled to enter and use the enlighten installations to find out more

Inside the exhibit, there is a clear indication of the extent of each Enlighten instance although the technology itself, of course, remains hidden. The back wall is subdivided into four vertically adjoining regions each separately titled and featuring, as centre pieces, single large representations of stellar imagery

(approximately one metre square) together with a number of smaller images and diagrams positioned along their bottom edges (Fig. 4.15). The centre pieces are chosen to represent, in stages between each installation, the sheer scale of resolution achievable with the SDSS Telescope, depicting images of first one star, then a million stars, and so forth. At each stage, exploration of these stellar phenomena using one of five available flashlights (twenty in total) reveals most of the educational material by utilising hidden or invisible targets. The aforementioned lower images, however, form obvious visible eye catchers and these provide further educational material on topics such as The Growth of the Universe, Orbiting Satellites, Spacewalks, Solar Eclipse's, Planets and Famous Scientists. Some sections of the display are high contrast (e.g. bright photographs against the dark sky background), requiring reasonably high dynamic range cameras.

Fig 4.15 – Working through the exhibit left to right, visitors can explore the universe at different resolutions starting with stellar photography of first “one star” (top left), then “a million stars” (top right), “a million, million stars” (middle left) and finally entire galaxies (middle right). Each “booth” features four additional photos or diagrams (bottom) designed to encourage interest in related subject areas.

The basic layout of each instance of Enlighten is found in the form of a tilted table, located between two pillars facing the associated wall section. These tables, as in the MRL installation, imply, but do not enforce a boundary between the torches and the back wall. While children do occasionally run through the main body of the exhibit, this is not common. The table itself houses a computer running the Enlighten software which is hidden from view by ventilated side panels, in order to allow for adequate air circulation within. On the surface of the table, various futuristic graphics and text explain how to begin using the exhibit, alongside five holsters containing maglights which are modified to be mains powered, via an extendable tether, thus connecting them to the bottom of their associated socket and preventing unauthorised removal (Fig. 4.17). Like the previously discussed

Enlighten installations, Explore the Universe uses audio for content delivery and this is achieved via two sets of directional speakers, set into the enclosing pillars, at both child and adult head height (Fig. 4.16). Due to the generally loud level of background noise found at Intech Science Centre, and the proximity of the adjacent installations, such a design choice was necessary both to make content audible to all those using the exhibit while at the same time avoiding causing interference to other visitors' experiences.

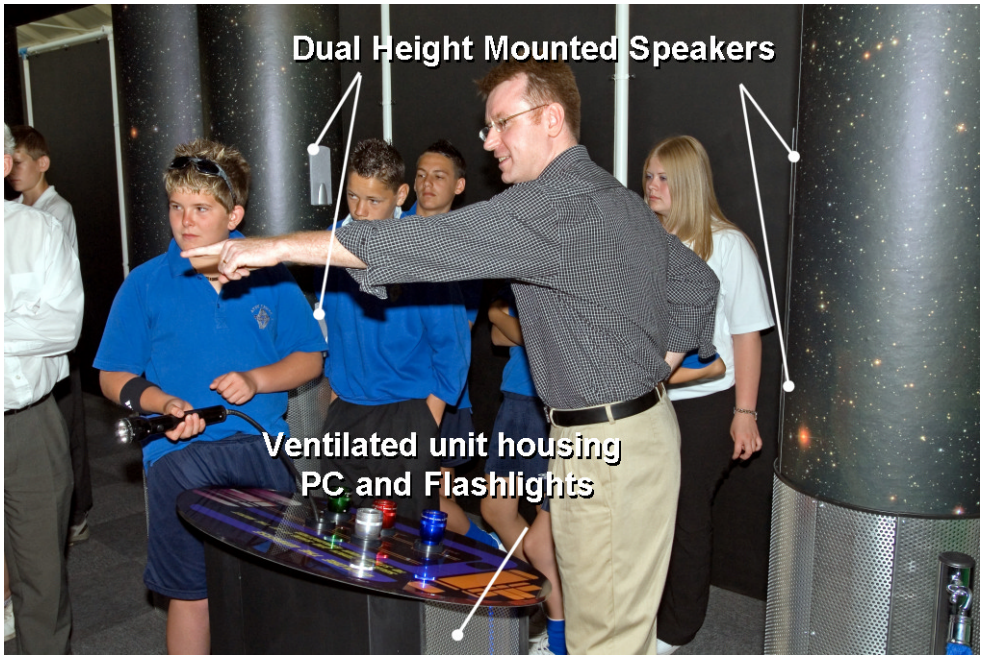


Fig 4.16 – Each Enlighten installation is self contained housing its own PC to run the software within the *torch table* that forms the centre piece of the display. Dual height speakers face inwards from pillars that imply separation between units but still allow space for small groups to interact with a single installation as a shared experience.

Unlike other installations such as Etruia, where use of a flashlight to trigger audio clips is not explained to visitors and is effectively something to "be discovered", at Intech, the intention for the public to make use of these devices is strongly advertised. This is done via the table in each booth, where flashlights are very prominently placed alongside information detailing their names and how visitors should use them to begin exploring the display (Fig. 4.17).

From an educational perspective, the flashlights to be found built into each of the four installations are categorised to be geared towards specific age groups. The

‘Basic’ and ‘Beginner Facts’ flashlights cover the key stage 3 and 4 curriculum with additional inspiring facts and information about the universe appropriate to this level of understanding. The ‘Advanced’ and ‘Further Facts’ flashlights cover GCSE level curriculum topics but also provide further details on the latest scientific observations and findings resulting from recent analysis of the most up to date data from the sky survey. Finally, there is also a more general, ‘Did you know?’ flashlight, in each installation, which provides less narrowly focused information about the universe and is designed to promote and encourage a broader interest in the wider subject area. For the purposes of display however, none of these categorisations are used for labelling and instead flashlights are given the contextual titles ‘Space Cadet’, ‘Mission Facts’, ‘Astronaut’, ‘How big how far?’ and ‘Galactic Commander’ respectively. This achieves an abstraction effective enough not to alienate the general public with an educationally specific naming convention but which teachers and leaders of school parties can be given prior warning of, in order to instruct their pupils as to where their attention would be best prioritised.

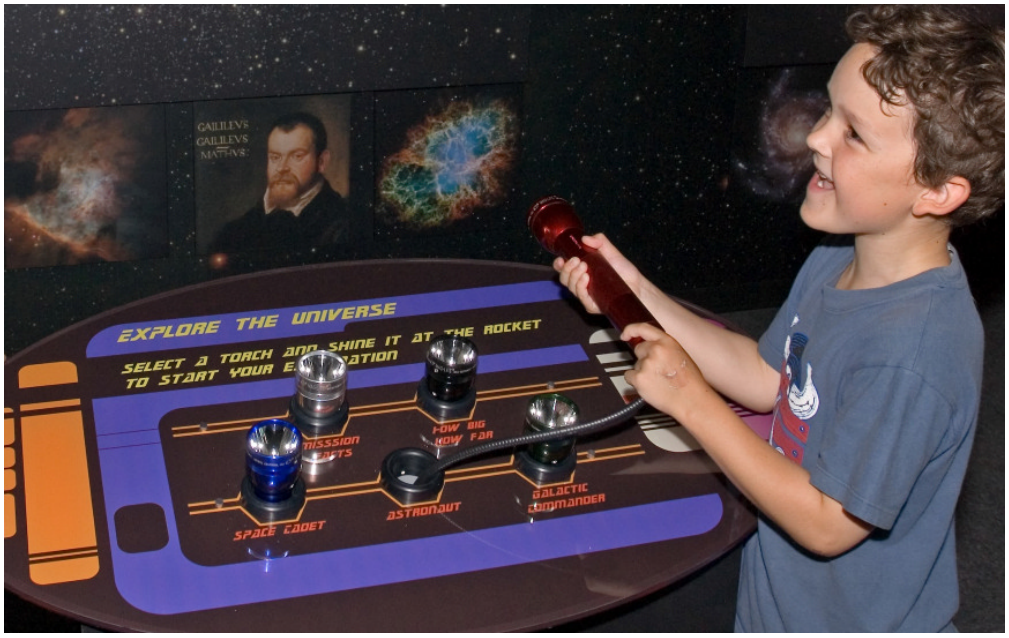


Fig 4.17 – Use of “Explore the Universe” is clearly labelled alongside abstract names for flashlights to indicate varying subject matter and target age range. Each installation also features a “start target” (a cartoon rocket) which, when triggered, further explains how to use the flashlights to interact with the display.

In the MRL installation it is possible, and indeed preferable, to make use of two flashlights simultaneously as the combination of audio narration heard against a backdrop of (mostly) classical music produces an engaging and pleasant effect with the additional advantages of encouraging collaboration. At Intech however, because the audio content consists entirely of clips containing spoken word, having more than one clip play at a time would make the narration hard to understand. It is for this reason therefore, that the Intech version of Enlighten has been customised to prevent this. When a visitor removes a flashlight from its socket this effectively disconnects all others connected to the same table until the original is replaced, meaning that it is only possible to ever hear one piece of audio content playing at a time, per installation. Because of the proximity of the nearby instances of Enlighten in Explore the Universe, activation of a flashlight also overrides the standard "identification via recognition" technique (Chapter 5) and, instead, this is achieved electronically using switches. Although such an alteration has obvious benefits as regards robustness and stability of the exhibit (no vision system having ever been truly one hundred percent accurate) the predominant reasoning for it is to prevent interference from visitors who might use an adjacent installation's flashlight either accidentally or deliberately on the wrong display area. This alteration not only prevents this from occurring whenever an installation is already 'actively' in use (audio clips are actually playing) but also ensures that, when it isn't, any potential interference can only trigger the correct selection of audio clips determined by its currently selected flashlight. When no flashlights are being used of course (they are all returned to their sockets) it is impossible for any interference to occur at all.

The modifications made to Enlighten in order to accommodate these changes were minimal and simply involved replacement of some of its event triggering logic (Chapter 5) by interpreting electronic signals (relating to each flashlight's on/off status as governed by whether or not it is docked in the table) delivered via the computer's standard RS232 port. In addition to these interface alterations, the software was also changed to incorporate an improved version of the ambient background sound functionality, used in the cave demonstration (see Chapter 2).

Although changes from the earlier version of this feature were primarily cosmetic (position governed cross fading between defined fore and background target regions), the addition of this functionality not only provides aural confirmation of the exhibits functional status, resulting in a more streamlined experience, but also an audible indication to visually impaired visitors that allows them to gain awareness regarding the boundaries of the active display area.

The identified technical issues relevant to this installation are:

- High relative contrasts between some target areas and the predominantly dark background imagery
- A potential for occlusion to be caused by members of the public obscuring the camera's field of view
- A potential requirement to recognise a large set of flashlights

4.5 MAGNA Science Adventure Centre

MAGNA Science Adventure Centre is the first of its kind to be created in the UK. Its unique location, housed within what was once Europe's largest and most prestigious steel works, sets it aside from many similar museums and science centres making it an intriguing and exciting destination for enthusiasts, local veterans, family excursions and educational field trips. Located in the beautiful countryside of South Yorkshire's Don Valley between the historic towns of Rotherham and Sheffield, the Magna Trust (a registered charity) and Millennium Commission funded centre was established in spring 2001. MAGNA plays host to over 400 thousand visitors per year and has recently been awarded the title of "Best Events venue in the UK". Like Intech, one of MAGNA's chief priorities is education. Special programmes and structured events run for schools all year round during term times. Primarily though, the focus of MAGNA is to provide a unique, interactive learning environment for everyone. To this end, 'hands on' exhibits have been designed, taking inspiration from the generations of steel

making at the site, around the classical elements of Air, Earth, Fire and Water. MAGNA creates an experience where science is aligned seamlessly beside art, design, technology and industrial history (Fig. 4.18). The Air Pavilion covers the forces of wind, vibrations that create sounds or shake buildings and also how pollution contributes to the destruction of the environment around us. The massive role that water plays in our every day lives and its uses both by, and to power machines is the topic of the Water Pavilion, while the two Earth and Fire Pavilions bring the visitor closer to the subject of steel making itself. Between them, insight is provided into the practicalities of mining, the machinery it requires and how fossil fuels are created, before finishing with exhibits regarding the heating, shaping and cooling of materials, energy transfer, recycling of scrap metal and the critical uses of both electricity and magnetism in all of these processes.

Fig 4.18 – MAGNA Science Adventure Centre. Themed interactive exhibits are structured around the classical elements of earth, air, wind, fire and water to provide an exciting, hands on experience (images © MAGNA)

In addition to the four ‘element pavilions’ and the two educationally themed outdoor playgrounds, Sci-Tek and Aqua-Tek, which guide children on a physical and fun exploration of forces, materials and how water treatment works, the Centre’s main and largest draw is, of course, the exhibition and history of the steel works itself. Starting with “The Face of Steel”, the UK’s largest existing multi-screen display featuring seven huge video projections, sound, special effects and a multitude of stairway mounted, smaller screens that depict imagery of heat, danger and community, visitors climb up to a gantry suspended a dizzying 150ft above the floor of the steelworks itself. Here they traverse an aerial walkway over half a kilometre of raw industrial history amidst the gloomy, harsh environment that was experienced by over ten thousand workers smelting, shaping and rolling steel in days gone by.

At the far end of the walkway (termed the “Sheffield End”) is found ‘E furnace’, the last of the steelwork’s huge electric arc furnaces which forms the focal point of ‘The Big Melt’ (see Fig. 4.19), a timed light, sound and pyrotechnics display (scheduled to run five times a day) where, movement, flames, sparks, smoke and authentic narration bring the furnace back to life. Despite such a dynamic exhibition at the culmination of the steelworks experience, the remainder of the walkway although located in an extremely content rich and stimulating area, previously presented little information. From the gantry here, visitors peer off into the vast expanses of the steelworks’ dark interior where gigantic hooks, crucibles, ladles and cranes (Fig. 4.20) lay silent and unexplained in amongst other long abandoned machinery. Due to the authentically dark setting, the position of the gantry and the lack of current augmentation in the area, the curators of Magna believed this to be the perfect environment where flashlights could be used naturally by visitors, as a means to go about discovering the secrets hidden within these, previously inaccessible, parts of the steel works.

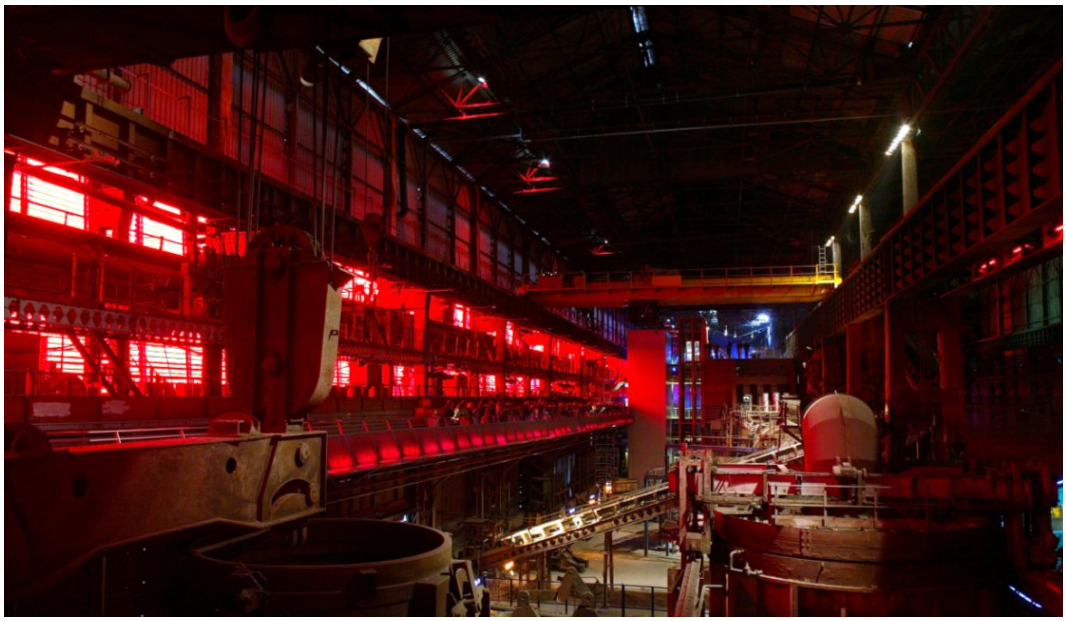


Fig 4.19 – The “Sheffield End” of the aerial walk way features “The Big Melt”, a light, sound and pyrotechnics show run five times a day. During inactive periods, a newly installed version on Enlighten allows visitors to interactively explore the various sections of “E Furnace” (right) from the safety of the viewing gantry (left).

Fig 4.20 – Exiting “The Face of Steel”, visitors cross the steelworks from above via an aerial walkway from which can be seen giant hooks, cranes and other machinery, perfect for discovering more about using flashlights.

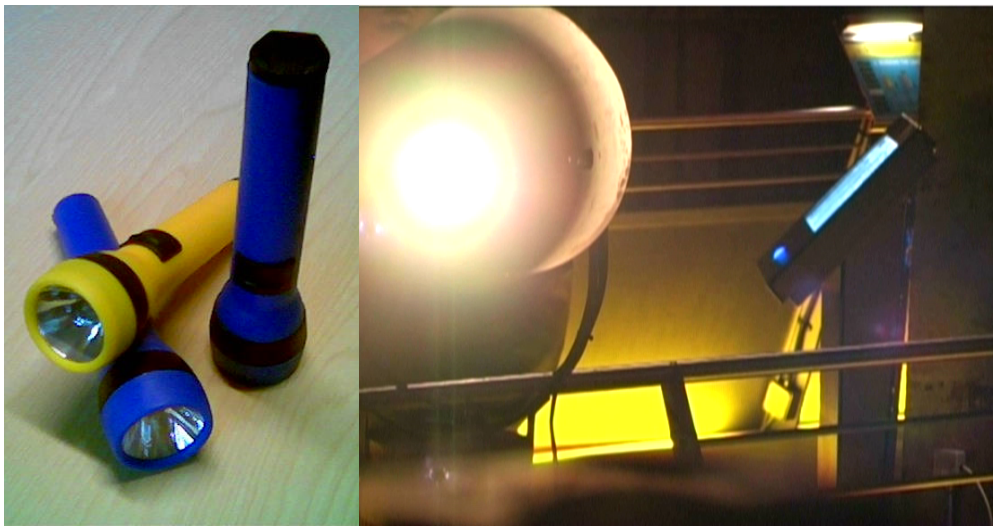
The proposal, put forward to, and eventually funded by, the National Endowment for Science, Technology and the Arts (NESTA) was to make use of and develop Enlighten within the steel heritage part of MAGNA over two phases, each

designed to test and stretch the technology in order to evaluate and learn about it's suitability within this, and other similar environments.

The initial phase of the project was to primarily scale up both the range and size of the deployment of Enlighten, as compared to previous installations, but to also extend and vary the type and design of the triggered content to involve more than its previous traditional narrative audio. Once this is completed, the secondary phase, concentrates on two areas, the development of the systems interface and the creation of a "ghost tour through the lower part of the steelworks" (see figure 4.23) where visitors make use of miniature "disposable" (i.e. cheap to manufacture) flashlights (Fig. 4.21i) to hear stories of, and hence learn about, the dangers of working with steelworks machinery in the past. Specifically, the aims of these two sub targets are; to develop a system that can be used portably and with ease by a museum exhibit designer (equipped only with a tripod, camera and laptop) and to extend the range and type of flashlight interactions available, making them capable of not only triggering, but also interacting with allocated media content. This may include, but is not limited to, recognition of overlapping flashlights, exploitation of flashlight gestures and interaction with moving film footage. Phase two is, at time of writing, still under development so a detailed description is beyond the scope of this thesis. For the remainder of this section we concentrate on stage one, the deployment of Enlighten on the previously discussed gantry, which forms the exhibit now known as "Steel Reveal".

Although with very powerful flashlights (top of the range MagLights etc) it is possible to illuminate parts of the building and machinery from the gantry, the effect is not dramatic. There are a number of additional problems with regard to deploying lights in such a fashion in museums. These range from issues such as mounting, cost and difficulty of customisation, durability (filament bulbs proving to be too fragile at Intech for example and instead needing to be replaced by LED's), charging/battery life and possible public theft. To combat this problem, and to aid with the scaling up of the system, MAGNA instead commissioned the manufacture of a number of large steerable spotlights (six in all, as in figure 4.21ii)

and these are placed at a number of locations along the walkway, each being used to reveal different forms of information or control special effects.



i) ii)

Fig 4.21 – i) the small, short range, "disposable" flashlights set for use on a ghost tour during the secondary phase of the MAGNA installation. ii) The massive spotlights installed at set locations along the steelworks gantry.

The first of the "Steel Reveal" spotlights is positioned at the start of the gantry, just past the end of the "Face of Steel" experience (see Fig. 4.23 for spotlight locations). Upon reading a short description of how to control the light to find information (found also on each subsequent spotlight), visitors seek out targets within the nearby areas of the building in order to learn about the life and experiences of steelworkers and their families in the work's "extended community". It was the intention of the MAGNA exhibit designers that such material would be projected onto a large screen, suspended nearby, allowing several people to share in the experience of the triggered content. This, due to cost and practicalities of rigging both screen and projector in suitable locations proved to be unfeasible. Instead, the first spotlight and the majority of subsequent ones make use of raised up "media stations" containing LCD displays and speakers capable of relaying a variety of different content forms (Fig. 4.22). The discovery of so called "content hotspots" with no. 1 spotlight now triggers archival footage to play on the screen. This includes "Spare Time", a short documentary film made in 1939 describing life in and around the steelworks during that period of history. Like those at Intech however, the curators of

MAGNA are fully utilising the reconfigurable nature of Enlighten and this, combined with the fact that the software has been modified to feed into the centre's generic MIDI control grid means content is easily changeable and, at time of writing, is still being altered in response to studies and visitor feedback in order to achieve the best experience possible.

Fig 4.22 – Children guide a giant spotlight to learn about life and conditions when working in the steelworks. Videos, audio and photographs are triggered for playback on a nearby "media station" controlled using flash animation.

Continuing along the walkway, the second spotlight is located on the opposite side to the first (Figure 4.23) and can be used to illuminate and hence learn about (through audio played back over a loud speaker system in this case) the complex machinery making up the secondary steelmaker that is located there. Challenges are provided here by the fact that there are almost no smooth surfaces within the target machinery. The system must be able to operate, when the spotlight beam is incident on a highly three dimensional surface, visible, sometimes, as only a collection of thin specularities arising from a sparse array of pipes.

The third spotlight on the Rotherham walkway is intended for visitors to learn more about the working environment within the steelworks. Here large, prominent features within the building can be illuminated such as the overhead cranes, giant hooks, ladles and crucibles which trigger playback of the deafening sounds that would have been experienced when the machinery was in operation. In addition to these, special effects such as smoke machines can be triggered, in order to seemingly bring the equipment back to life, as well as a number of less obvious "hidden targets" (again following on from their successful use within the Nottingham Caves) to encourage a sense of discovery. Finally, before entering into "The Big Melt" area, use of the last spotlight triggers a collection of silent slideshows (displayed on a nearby media station) depicting family generations of steelworkers, formed out of the community, through the years.

Despite the "Sheffield End" of the walkway already being highly augmented with the scheduled "Big Melt" show (Fig. 4.19) it remains an extremely feature rich area, ripe for discovery and learning using Enlighten. Here, use of the spotlight (which is only active outside of scheduled run times) covers material related to that contained within the show. Visitors can explore and develop an understanding of the various stages required to operate 'E-Furnace' and this is conveyed by use of Flash animation which is displayed on a nearby media station. The final spotlight in the area (also disabled during show times) is used against a photographic display of archival images of steelworkers which come to life when illuminated, gaining the visitor valuable anecdotal experience of the accents, dialects and songs which represented an important and prominent part of daily life for them. In all, visitors can experience up to six hours of operating time with these final two spotlights, and an hour of the Big Melt demonstration per day.

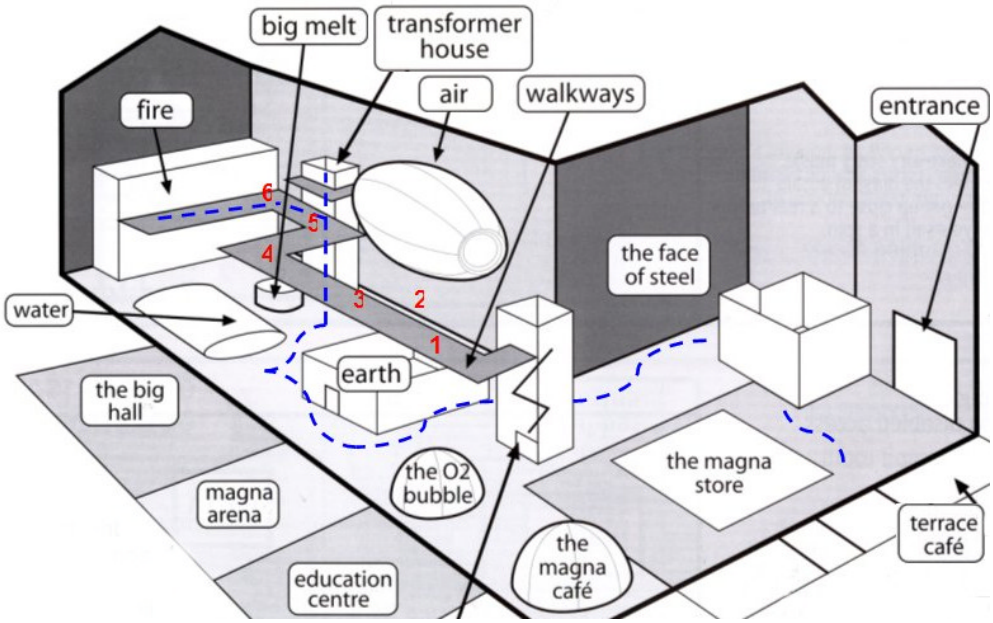


Fig 4.23 – Layout and positions of Enlighten spotlight installations at MAGNA as described. Key: 1 – The steelworks community, 2 – The secondary steelmaker, 3 – Working environment, 4 – Generations of steelworkers, 5 – Exploring 'E-Furnace', 6 – Accents, dialects and songs of the steelworks. The dashed blue line indicates the proposed route for the "ghost tour" which will form the basis of the secondary phase of the project.

The Enlighten installation at MAGNA, which has been open to the public and in daily use since October 2006, is the most ambitious to date. This is due to the deliberate pushing of the systems operational ability, and to a number of environmental issues and practical difficulties with regards to installation.

Like Etruria, the steelworks environment is in many ways an outdoors location (with the exception that it is fully enclosed and devoid of natural light) requiring the use of toughened/protected durable equipment to withstand the effects of dust and temperature variations, with additional consideration of factors such as the thickness of coax cable used, in order to avoid damage from rodent infestations. Although resilience to precipitation is not a factor this time, the inside of the steelworks' temperatures are often lower than those outside. Systems must therefore be configured to allow for possible visual features, such as illumination of spotlight beams (rather than just the location where they intersect with surfaces) as might similarly be observed from standard flashlights when used on a cold misty night. Such illumination features are also due in part to the low levels of light available within the building. These, additionally, can present difficulties

with the acquisition of reliable background imagery that, as discussed in chapter 5, forms an indispensable part of the image processing algorithm.

The MAGNA installation also presented a number of practical issues. These ranged from the significant amplification of speaker equipment (in order for audio stations to be heard against excessive background noise) to the obvious difficulties regarding mounting of cameras under a gantry suspended 150 feet above the ground. Centralised installation of computer equipment (in a control room positioned far away from actual exhibitions) also presented a challenge. Aside from the physical limitations of cable runs (potential degradation of signal strength), this effectively required non-interactive configuration of each area, or configuration using two people communicating via CB radio. As in Etruria and Intech, the MAGNA installations of Enlighten also required the inclusion of an intuitive interface and the ability for systems to be 100% automatic when starting up, once initial configuration has occurred. The alteration of system logic in order to fire the aforementioned MIDI signals in response to target triggering was the only significant change from these versions.

The identified technical issues relevant to this installation are:

- Potential variations in spotlight beam appearance under cold conditions
- Very Dark environment: potentially causing noisy representation of background imagery or a system that is less responsive to high speed movement (if frame rate is decreased in order to counteract low light levels)
- Sparse, three dimensional or highly reflective, targets

4.6 Conclusion

The feedback, ideas and enthusiasm provided by staff at MAGNA and the other discussed museums presents further evidence to support the suggestions made in

chapter 1 regarding the application of flashlights as useful, practical and engaging interaction devices and their suitability for augmentation/creation of public exhibitions. There are, however, some technical issues that have been identified as being relevant to the above described installations and these must be resolved in order for such a system to be maximally effective.

4.7 Summary

The work reported in Chapters 1-3 suggested that interactive devices based upon visual tracking of torch beam are potentially useful and technically feasible. Such devices are, therefore, likely to find application in a wide variety of situations and environments, as evidenced by the installations discussed above. Five environments for such a system have been described in full, each description detailing information about the location of the installation, the exhibit augmentation proposal, the constraints of deploying a system in this particular location, a summary of the environmental factors under which the system must operate and any special configuration/system interface requirements required. Together with the above, a number of key technical issues have been identified. These are, in some cases, relevant to more than one installation and therefore very likely to affect future installations.

In the remainder of this thesis we examine the various components and stages of processing that build up or are incorporated into the above described systems. Attention is focused on how the identified technical issues can be dealt with. A detailed discussion of the development, optimisations and difficulties relevant to the implementation of each component is also provided.

Chapter 5

Design Issues and Decisions for an Improved Flashlight System

The interactive flashlight systems described in Chapter 2 employed standard (moving) object detection techniques. In Chapter 3 a method was developed which allows information about the pattern of light projected by a flashlight to be extracted. In what follows we describe an improved interactive flashlight system based upon the new operator. This system, Enlighten, has been deployed in the environments and circumstances discussed in Chapter 4. That deployment identified a number of technical issues that had to be overcome.

We now examine Enlighten in detail, discussing its design, features, optimisations and the difficulties that were encountered during development of its various components. To this end, we first provide a system overview and then look in detail at its chain of processing, split up into five sections covering background estimation, operator application and optimisation, threshold techniques, training and recognition, and media triggering logic. Descriptions of some, as yet, un-resolvable issues and discussion of their potential (but impractical to implement) solutions is provided in Chapter 6.

5.1 System Overview

Enlighten's main processing is driven by the frame rate of its connected cameras (typically 30fps) and, in order to function, applies a chain of separate processing steps to extract required data and trigger events. Conceptually there are effectively only two separate modes governing which processing stages should be completed.

These are:

1. Those applied when training
2. Those applied when the system is actively seeking to recognise flashlights.

In fact, there is in reality very little difference between the two modes as, when recognising flashlights, candidate regions must have features extracted in the exact same manner in which they are extracted during a training phase. The main difference lies simply in whether or not such features are stored as recognisable elements for an individual flashlight, or instead compared to all such features stored in memory, in order to find the best candidate match. Since the differences between modes are minor, graphically the flow of processing is best represented by a single stream. Although the remaining sections of this chapter consider these various stages in detail figure 5.1 presents a brief overview.

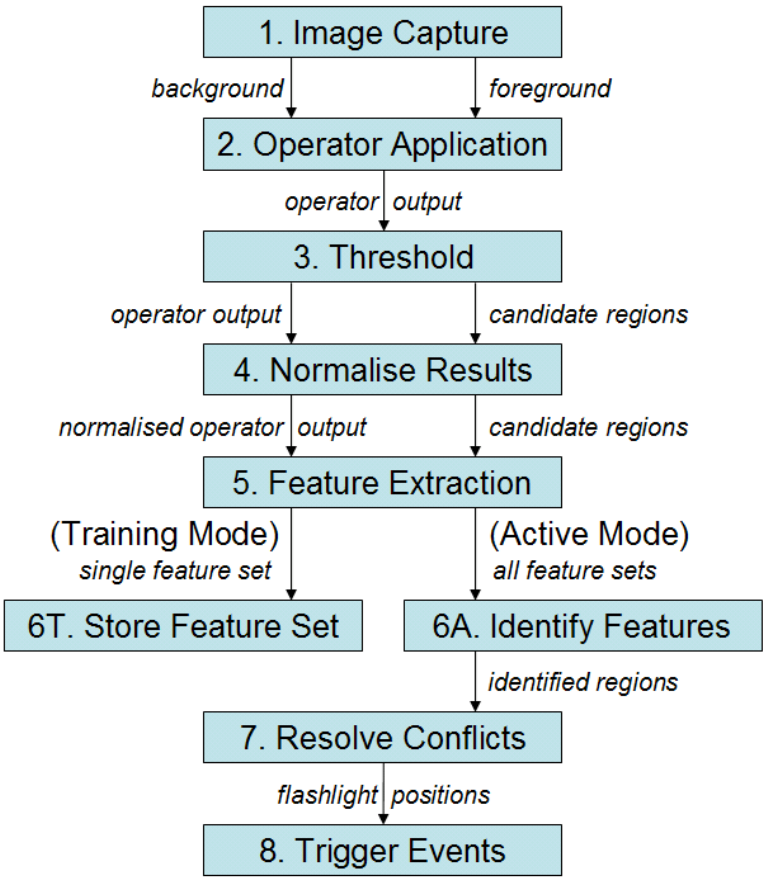


Fig 5.1 – Processing Flow Overview. 1. Initial (at system start-up) capturing of background images allows a representation of a scene to be constructed. Subsequent images are used to update and maintain this representation. 2. using fore and background image pairs a representation of the scene featuring only illumination unaffected by surface reflectance is constructed. 3. A Dynamic threshold technique is applied to this representation to identify regions most likely to contain flashlights. 4. Data associated with candidate regions is normalised to remove the potential effects of long-term variations in ambient illumination 5. Features are extracted from candidate regions (only one region is permitted during training). 6T. (training mode) feature set is stored to construct profile of an individual flashlight. 6A. (active mode) feature sets are compared to those from training data held in memory to determine the most likely identification for each. 7. Conflicts between competing regions are resolved 8. Using positional information, events are triggered such as audio/video playback or special effects

5.2 Background Estimation

5.2.1 Motivation and Issues

As discussed in Chapter 3, the Quotient operator which forms the basis of recognition within the Enlighten system, requires there to be available, at all times,

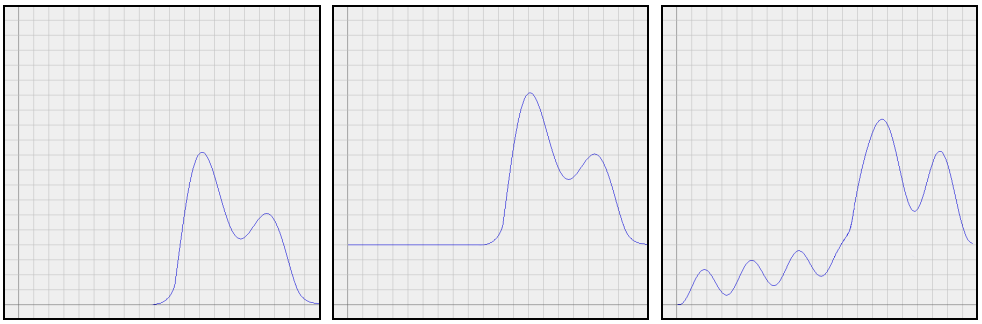
both a fore and background image pair. The results presented (in Chapter 3) suggest that use of just a single frame to represent the background image (the monitored area when devoid of additional incident light from a flashlight beam) provides good results and is of course trivial to implement. However, despite the success of this, even simple averaging of a few frames produced even better results in the experiment reported in Chapter 3.

Though obviously desirable, improved representations of flashlight projections per se are not the only motivation here. In Chapter 4, for each experimental environment discussed, a summary of the main identifiable technical issues relevant to that installation are provided. It is notable here that three out of the five documented were effected by changes in the level of ambient illumination over time. Problematically, this typically led to reduced consistency in flashlight recognition, false triggering in the absence of a flashlight beam actually being present or failure of detection altogether. As described, the three afflicted installations (The Lab, The Newark Show and Etruria), regardless of their indoor locations, were strongly influenced by natural light and it is therefore likely that any future semi-enclosed or outdoor installations would be similarly effected.

Considering the above, it would appear that maintaining an up to date background representation is mandatory for consistent system performance. However, despite this indication, examination of our operator reveals that, in theory, variations in ambient illumination should not in fact have any influence that cannot be accounted for. Therefore background maintenance should not actually be required.

To explain this hypothesis, consider that the base assumption of our system, that ambient illumination should not alter over time, has of course been violated. However, should our secondary assumption, that ambient illumination remains spatially uniform, continue to hold, it should still be possible for our threshold algorithm (see section 5.4) to identify regions containing flashlight beams, irrespective of any changes. This is because, under our assumptions of uniformity, any changes in such illumination will appear in our processed results as if we have

turned on (or off) a large (existing over the entire field of view of the camera) flat and featureless floodlight. Given that such light presents a vastly different profile to that of a typical flashlight (Fig. 5.2), the threshold algorithm should be able to disregard it.



i) ii) iii)

Fig 5.2 – Cross sections through the centre of a flashlight profile as represented by operator results. Graphs i) through iii) demonstrate effects of absent, spatially uniform and spatially varied changes in ambient illumination. Adaptive thresholding can reasonably be expected to isolate the flashlight projection in i) and ii), though iii) may be problematic.

The manifestation of ambient illumination changes in results (Fig. 5.2) can also be demonstrated mathematically by factoring in a second unknown $I_{\Delta A}$ to represent them in calculations as shown. This is different of course from the initial level of illumination I_A present when the original background representation is constructed. Equations 5.1 and 5.2 give familiar definitions of intensities recorded in the background and foreground images respectively. Note the inclusion of $I_{\Delta A}$ in eqn. 5.2.

$$e_B = I_A R \tag{5.1}$$

$$e_F = (I_A + I_{\Delta A} + I_T)R \tag{5.2}$$

Substituting eqn. 5.1 into 5.2 gives

$$e_F = (I_A + I_{\Delta A} + I_T)\frac{e_B}{I_A} \tag{5.3}$$

which can be rearranged to show the effect of $I_{\Delta A}$ on the output of the Quotient operator.

$$e_F = e_B + \frac{I_{\Delta A} e_B}{I_A} + \frac{I_T e_B}{I_A} \tag{5.4}$$

$$\frac{e_F}{e_B} = 1 + \frac{I_{\Delta A}}{I_A} + \frac{I_T}{I_A} \tag{5.5}$$

$$\frac{e_F}{e_B} = 1 + (I_{\Delta A} + I_T)\frac{1}{I_A} \tag{5.6}$$

$$\left(\frac{e_F}{e_B} - 1\right) I_A = I_{\Delta A} + I_T \tag{5.7}$$

Under experimental conditions we experience near identical ambient illumination levels between capture of fore and background imagery. When changes in levels ($I_{\Delta A}$) do occur (eqn. 5.2) between capture of these two images however, these are shown to be simply added to the output result (eqn. 5.7).

5.2.2 Deploying Adaptive Background Estimation in Enlighten

When considering the classical problem of capturing reliable background representations, which are devoid of the content we wish to detect (flashlights), the immediately apparent solution would simply be to make use of our detection mechanism to determine when flashlights are present, and update the background representation whenever this isn't so. Unfortunately, such a method is flawed by the fact that flashlights are rarely merely "there" or "not there" but instead appear gradually over a few frames if ignited "in shot" (see section 5.8) or are found to be only partially in view, for a period, if entering from a side. Such circumstances typically lead to a scenario in which sections of a flashlight, or elements of a partially ignited one, would contribute to background estimation and this in turn would lead to degradation in recognition, over those areas of the background representation containing false data. In addition to this, the method relies on our detection technique being 100% accurate (which few vision techniques are) meaning that a falsely detected flashlight in a single frame could lead to background estimation being permanently turned off.

Instead a technique is required that works independently of the detection component. One explored avenue lay in analysis of the changes in image intensity to determine if simple statistical measures could be used to identify the nature of the variation in illumination between capture of a fore and background image pair.

If we assume that changes in ambient illumination are spatially constant across the field of view, they will increase operator output by a constant (Fig 5.2ii) and the mean operator intensity will therefore also increase by the same amount. The standard deviation of the operator output will, however, remain the same. In contrast, a flashlight, which under normal conditions, one would expect to have a smaller effect on the average operator output, should exhibit a more pronounced effect on its standard deviation. The possibility arises of differentiating between changes in ambient illumination (sizeable change in mean operator output, constant standard deviation) and the addition of a flashlight (smaller change in mean operator response, larger change in standard deviation). If this variation could be reliably identified, it might be used as a means to determine when it is safe to utilise the current foreground capture, to either update or contribute to the current background estimation. In practice however, the differences in operator statistics between those gathered in both presence and absence of a flashlight are too similar to be reliably separated from those statistical differences observed due to small ambient illumination variations.

The method chosen for application in Enlighten instead fell back on a more established technique making use of median filtering (e.g. Grimson et al 1998) over time. This features the well documented advantage of being able to utilise all (if necessary) captured images, regardless of their content, so long as features that are not part of the actual background do not remain stationary, and in shot, for the window of frames over which the systems' background image is calculated. This technique, for the majority of the scenarios discussed in Chapter 4, is particularly suited to a flashlight interface. As discussed in Chapter 2 it is typical behaviour for beams to travel fast and in a seemingly random manner. One restriction that is applied however is that, due to the length of audio messages (e.g. those at Etruria)

used within some of the installations, the time frame (or window of frames over which the background image is calculated) had to be significantly extended. This was to avoid the stationary flashlight beam of a listening user being effectively "burned into" the background estimation. Although, in these cases, such customisation makes the system less responsive to particularly rapid changes in ambient illumination, for the most part, this has not been significantly detrimental to overall system performance.

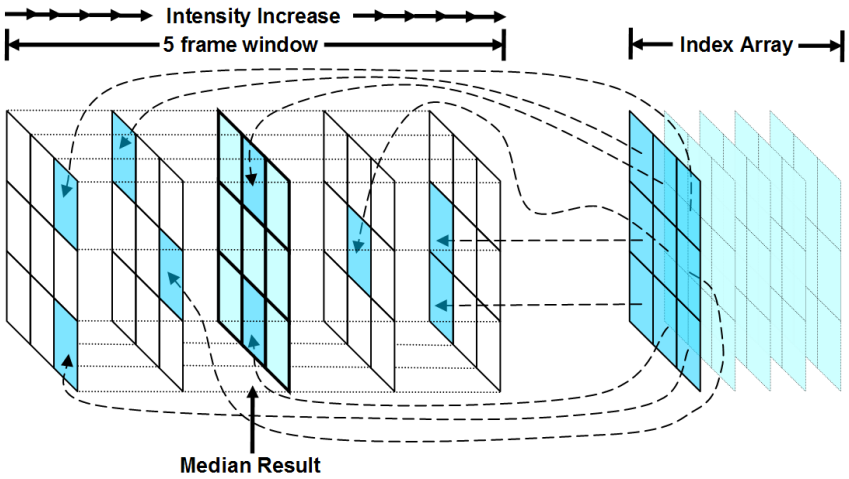


Fig. 5.3 – Median Filtering: Video is captured directly into an *n* framed processing "window" with each frame's individual pixels ordered separately along the z-axis (according to its intensity) and an index array utilised to record their distribution. After each update, the centre frame is used to represent the new background estimation then the oldest frame data is removed and replaced with that from the next update.

As usual, the method used consists of collecting a number of individual frames (window), ordering them and then picking the centre result as detailed in figure 5.3. In the case of Enlighten however, it is also necessary that the *first* obtained background estimation be retained, as this forms requisite data, for part of a normalisation process, discussed later in section 5.5.1. Additionally, in real situations, it is important that Enlighten supports fully automatic start up from a power off state. This means that it is possible, or indeed highly likely, that cameras have been off for some hours and therefore take a moment to transmit a valid signal when first activated. To accommodate this, in addition to first loading configurations, previously captured background images and flashlight training data, Enlighten also delays normal processing for a short period. This avoids capturing any erroneous data from cameras that might otherwise contribute

to a false initial estimate of the background in a scene, effectively ensuring that any, flares, blank images or signal spikes, as are common on start up, are not included in calculations. Although such data would be neutralised in time thanks to the nature of the background estimation technique, this may not occur for some minutes due to the settings applied in some of the installations.

Due to the speed of the changes in illumination experienced in our experimental environments, it was not necessary (or indeed desirable due to processing overheads) to update the background estimation every frame. Instead, base settings were customised in each location from an initial configuration of one update every 2000 frames (80 seconds) meaning that, in a twelve frame window (as was used by default) a large sudden change in a scene would need to persist for up to eight minutes before manifesting itself in the background estimation. This allowed flashlights to be actively used for long periods without dire effect.

As previously discussed, for some locations, such a delay was unacceptable due to commonly rapid changes in ambient light. In these cases, shorter periods (taking note of how long flashlights are held stationary while a user experiences content) were used. However in Etruria, most of the ambient light entered from the kiln's doorway. The arrival or departure of a visitor therefore caused most, and the largest, changes in background illumination. As a result, a longer period was required, that was based upon a visitor's average stay inside (blocking the ambient light) rather than how long they were likely to use a flashlight.

In summary, adaptive background estimation is required to allow interactive flashlights to be deployed in real situations. The classic median filtering approach proved suitable, but care must be taken when choosing the time period over which the median is computed. In particular, the time period must be short enough to accommodate natural changes in ambient illumination, but long enough to accommodate the length of time flashlights are likely to be held stationary. In situations where the user's position affects background illumination, this must

also be taken into account. In practise a compromise between these requirements must be reached.

The median filtering approach proved effective in the set of real world installations considered here. It was noted however, that despite having a good quality background image, in several cases the output of the Quotient operator was clearly not completely independent of the reflectance patterns present on the target surface. This problem is discussed further in Chapter 6.

5.3 Pre-Processing and Optimisation

After the initial acquisition of background images, the first stage of processing within the system is the application of the Quotient operator (as discussed in detail in Chapter 3). This presents a representation of the scene from which the effects of background reflectance variations have been removed, thus allowing us to attempt recognition of any features remaining which might pertain to beams of light incident from a flashlight.

Following eqn. 3.20 (the Quotient operator), each foreground value is divided by the value of its equivalent background pixel (see below for the special case of calculations with a zero valued background pixel), and has 1 subtracted from it, effectively allowing for the representation of decreases as well as increases in illumination. As noted in Chapter 3, it is also necessary for the values calculated to be scaled by the level of ambient illumination present. This value is assumed to remain a constant throughout. Because, as discussed in Chapter 3, it is impossible to know this constant (without other knowledge such as reflectance properties of the viewed surface, for example) ambient illumination is universally assumed to be 1 and, consequently, this stage is omitted. This is shown functionally in eqn. 5.8 where i and j represent image coordinates in both fore (F) and background (B) image pairs.

$$f(i,j) = \frac{e_{Fij}}{e_{Bij}} - 1 \tag{5.8}$$

Some pre-processing is required before the operator can be applied. The fore and background images are acquired, converted to grey scale and filtered to remove noise, hence attaining the best results (as concluded in Chapter 3).

The derivation of the Quotient operator is based on the assumption that the viewed surface is a smooth Lambertian reflector (Lambert 1760). In many situations this is not the case. At the Newark Show, the marquee framework comprised a number of polished metal beams which introduced local specular reflections. Wet pebbles on the floor of the Etruria Flint Kiln produced a similar effect. In other areas of Etruria and at MAGNA the viewed surfaces were not always smooth, but comprised three dimensional, self occluding targets viewed against non planar backgrounds. In these circumstances the core assumptions of the method are violated and interaction cannot be supported.

To avoid this problem, pathological regions are removed by a logical AND with a user defined mask image. Masks are created using an art-package paintbrush style interface. The operator described above is then applied.

Because of the computational expense of convolved division on a frame by frame basis the method will only achieve full frame rate on a Pentium 4 specification machine typically using an image resolution of 320x240 pixels from a video stream. When considering the discrete nature of the image representation acquired from most webcams or standard video input frame grabbers, however, it is possible to see that such expense can be avoided. Given equation 5.8, it is clear that, under ideal circumstances, the range of potential output values gained from use of the Quotient operator is infinite. It is, however, of significant value to note that, when using a finite number of possible input values (as is the case with most digital, or at least consumer level, cameras) there are only a set number of calculations that ever need to be performed. Taking this into consideration, it has

been possible to re-implement the per-pixel runtime calculations described above, by making reference to a lookup table where computations for all potential combinations of input values to the equation have already been stored.

For the most part, each of the 65536 potential values (assuming input intensity ranges of 0-255 discrete levels) are calculated as described above. However, special care is needed when considering the circumstances where background intensity is measured as zero due to the infinitely high values such computations should in theory produce. Under ideal circumstances of course, a correctly configured system should be calibrated such that a good visible background representation exists at all times with, the somewhat ambiguous, level zero intensity (effectively meaning "fairly dark compared to everything else present in shot") never actually occurring. Unfortunately, due to circumstances discussed further in Chapter 6, together with potentially large decreases in levels of ambient illumination such calibration is in practice quite hard to achieve. Instead of effectively ignoring these abnormalities by setting their associated calculated values equal to zero, a different solution is required.

When both fore and background pixel intensity is zero the Quotient operator output can also be set to 0, as it is impossible for illumination to have increased in this scenario (at least within the sampling ability of our cameras and therefore not significantly). Where the foreground intensity is greater than 0, however, it is best practice to perform calculations with a pseudo background level of 1. Although this will not provide an optimal estimation for correct operator output, at this position, it represents the closest usable potential data point to the true value and the calculated result is less likely to appear as anomalous as it would have had it been set to zero. This is better illustrated in figure 5.4.

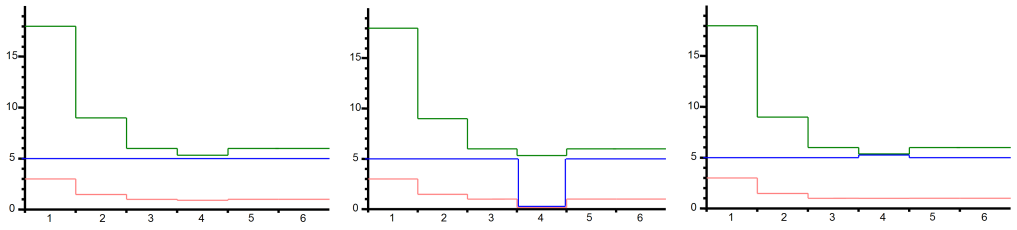


Fig 5.4 – Intensity representations of six adjacent pixels from corresponding background (red) and foreground (green) images where uniform illumination has been added to a surface that features varying reflectance levels. Assuming non-discrete sampling i) it is possible to measure fractional intensities (as found at pixel 4) hence a correct representation of added illumination (blue) is attained. Since cameras only offer discrete values, however, fractional intensities are measured as zero and this makes calculations unworkable. In the example shown ii), illumination at pixel 4 can be assumed zero presenting an inaccuracy of -5. An improvement to this technique however is to assume a background intensity measurement of one iii) giving (in this case) an illumination measurement of 5.4 that is only +0.4 from the correct value.

Utilization of the lookup table, as expected, realises a massive performance increase resulting in Enlighten being able to run on as low a specification PC as a Pentium 2 and therefore, when running on higher spec PCs, allowing the system to handle multiple cameras at full TV resolution (768x576) or run other processes at the same time (for example the flash multimedia display programs utilised at MAGNA). With today's commercially available multiprocessor and multi-core PCs there is scope for further enhancements however to date, due to the exceptional speed increase already attained by use of the above described optimisation, implementation of such functionality has not yet been required.

5.4 Thresholding the Operator Output

Once the above described process of applying the optimised Quotient operator and any potential user defined masks has been completed the output data needs to have a threshold applied so as to locate those regions within each frame that might contain a flashlight beam. This allows these to be further processed in order to recognise them (see section 5.5.2), or alternatively determine that no such regions exist and that the system is therefore inactive.

The simplest thresholding technique is to apply a user defined, global fixed threshold. Such thresholds are notoriously hard to determine however and this is also true when they are used for flashlight detection. This is demonstrated by the difficulties experienced with the technique detailed in Chapter 2. In addition to such difficulties, the use of fixed thresholds must also be ruled out, not only to combat factors such as varying illumination (see Chapter 4), but also to allow for a variety of flashlights to be identified. These may vary in both size and intensity.

Many adaptive thresholding schemes have been proposed (Sahoo 1988, Glasby 1993), with several generating threshold values as a function of some properties of a histogram of image intensity. A common approach is to assume that the histogram is bimodal, with each mode representing a distinct image region, and that the goal is to separate the modes (Kittler and Illingworth 1986). In a surprising number of situations, however, one of the modes is much smaller than the other, to the extent that the histogram is effectively unimodal (Joseph 1989).

Initial examination of the operator output revealed a high level of uniformity in the (low-valued) background, with flashlights appearing as compact regions containing higher values. Bimodal histograms were predicted as a result, featuring a large peak most frequently clustered around the zero mark and a smaller, but still discernible, one in the bins associated with higher output values. These would represent noise and flashlight distributions respectively. Finding a threshold in this case, could be a simple matter of splitting the two distributions by locating the trough in the graph that represents their histogram. In order to do this, to avoid false positives caused by small variations, it is first necessary to smooth the histogram by averaging its values over a local region. Once complete, the smoothed graph can be differentiated and a search for zero crossings should reveal a suitable threshold. In this case a smoothing window equivalent to 5% of the total number of bins was found to be effective in removing noise from the histogram.

Unfortunately, closer inspection of typical data reveals that, in fact, bimodal distributions are extremely rare in our scenario. This is because, typically, the

footprint of most flashlights' beams in a frame (as a percentage of total pixels) is relatively rather small. In practice therefore, in order to be discernable, and thus create the desired second mode of distribution, any beams found present would need to:

1. exhibit an almost completely uniform light distribution
2. feature a very narrow range of intensities (few histogram bins)

Uniformity of intensities in flashlights is, of course, unusual and, in fact, it is a requirement that flashlights are not uniform, in order for them to be recognised at all (see section 5.5.2). If such a phenomenon were to occur commonly this would leave size and/or shape as the only remaining features from which identity classifications could possibly be derived.

When observing real histograms of operator data containing flashlights, what we instead find is that a typical distribution, as expected, does indeed feature a large peak clustered approximately around the zero mark but that this distribution then drops sharply to depict low frequencies of intensity groupings. These commonly decrease within those bins representing the higher intensities (see figure 5.5). Effectively then, our typical results, when presented as a histogram, form a unimodal distribution. There is no clear separation, therefore, between noise, found in non-flashlight regions of a frame, and pixels that contribute to the light distribution within a flashlight beam's profile. Analysis of these results reveals an explanation for this. Although flashlight beams, incident on surfaces, can be perceived by humans to be significantly brighter than their surroundings, they do, in fact, always incorporate a gradient around their edges. This runs outwards from the perceived edge of the beam, down to the intensity levels of its immediate surroundings so therefore incorporates the most pixels. Typically, the brightest parts of a flashlight beam are found towards the central area. When considering distributions to be roughly circles, it is obvious that these regions, that contain the brightest intensities, also have the smallest circumferences. They therefore exhibit low frequencies in the histogram representation.

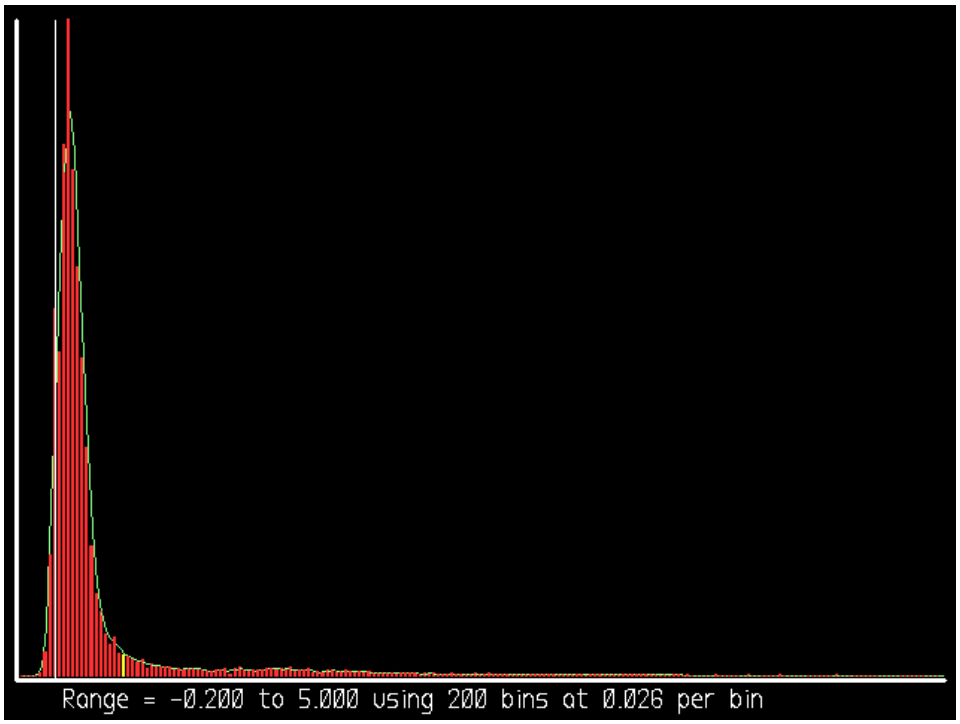


Fig 5.5 – Histogram representation of operator results

In order to apply a threshold, we need a technique capable of splitting a unimodal distribution at the divide between flashlight projections and background. This must also allow the extent of the regions, found to be representing flashlight beams, to be similar to those perceived by humans. As noted, flashlights, in actual fact, illuminate regions far greater than what most would consider the extent of their beam. It is important, from a usability point of view, however, that regions, created by a threshold, draw a parallel with the perceptions of the system's user.

One possibility, for determining such a threshold, lay in the technique of smoothing the data by averaging bin values over a small local window and marking the point of greatest change in frequency. This, however, commonly produced regions around flashlight beams that incorporated far too much, of what a user considered to be, background. Strictly speaking, the regions, produced by such a threshold, contained low levels of illumination resulting directly from the presence of a flashlight and this means they were correct. Since such a "correct" solution does not yield regions that tally with a users' perception of a flashlight beam's extent however, (see Fig. 5.6) a different unimodal thresholding method is required to achieve the desired results.

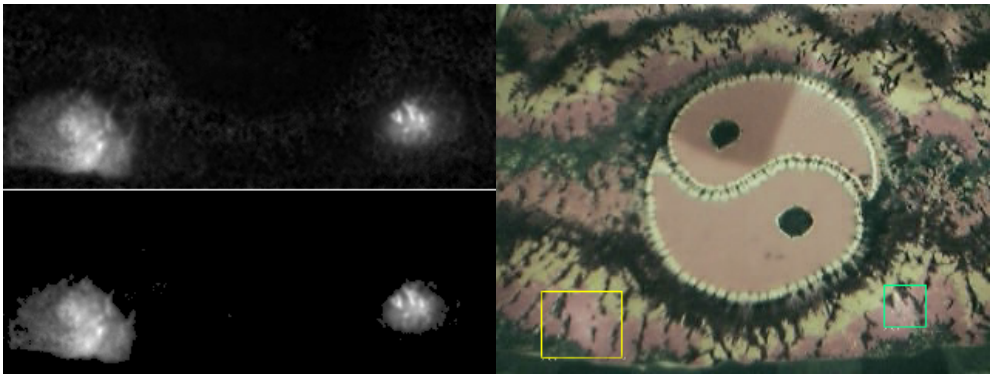


Fig 5.6 – The operator output (top left, scaled for clarity) is extreme sensitive to light making it possible, using the above described technique (bottom left, again scaled for clarity), to determine more of a beam's extent than a human might. This however may not be in line with a user's perception of the extent of a beam (marked right).

The subject of unimodal thresholding is addressed by Dunn and Joseph (1988), specifically for the case of processing poor quality line drawings, a problem that, due its similar ratios of a small amount of desirable data (the lines) to a large amount of noise, appears to have a lot in common with our described situation. Recognising the similar problem of lines creating a non-discernable peak in a histogram, Dunn theorises that the previously described noise peak may be considered to be a normal distribution. If so, by using its histogram representation, smoothed to reduce error, it is possible to identify a point along the x-axis, found at half the height of the distribution's peak. This corresponds to approximately 1.2 standard deviations (σ). Measuring this value from the x-coordinate corresponding to the peak, and multiplying it by three, can provide a position, either side of the peak, representing 3.6 standard deviations from the mode. Dunn finds such values represent markers that wholly encompass the noise distribution. In our case, any data above the rightmost of these could be considered to be resulting from a flashlight.

Under ideal circumstances, our implementation of Dunn's technique produced reasonable results. The global illumination changes experienced in the experimental installations (see Chapter 4) however, often reduced the consistency of the threshold levels applied. Thresholds would vary considerably in a very short space of time, disrupting interaction. Some of the reasons for this, as

expected, can be attributed to the Quotient operator occasionally producing results which are not entirely independent of the target surface's reflectance properties. This is addressed in detail in Chapter 6. However, more specifically in this case, problems occur due to the varying nature of the noise distribution when exposed to large changes in global illumination (Fig 5.7ii). As expected, dynamic background estimation (section 5.2.2) goes some way to minimise these effects over time and hence keeps noise distributions roughly around the zero mark. It has been observed however; the width of the mode is very sensitive to noise. When background illumination changes dramatically the half-width, and so the threshold value selected, becomes unstable. This is illustrated in figure 5.7i. For example, when background estimation is very up to date, it is common for a potential ninety percent of data to fall within the bins closest to zero. The mode effectively becomes a spike and it is very difficult to accurately estimate the width of the distribution.

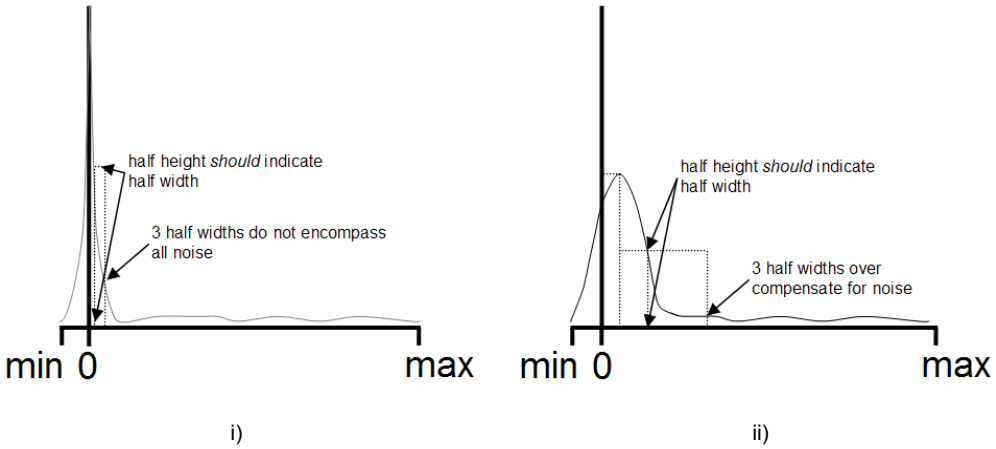


Fig 5.7 – Representations of histograms of operator output used for calculating thresholds via the Dunn and Joseph technique. i) Utilising very recent/accurate background representations, noise distribution is commonly too narrow for accurate measurement hence the assumption that noise falls within 3 half widths is incorrect leading to excessively low estimations for threshold values. ii) Under large illumination changes, noise distribution widens significantly as inaccurate sampling has more effect. In this situation, 3 half widths sets the threshold too high.

An alternative approach to the problem of thresholding uni-modal data sets is proposed by Rosin (1999, 2001). Rosin notes that, while bimodal threshold algorithms are commonplace, less attention has been paid to unimodal thresholding which in fact is a commonly found phenomenon. Such a

phenomenon often occurs in applications utilising edge detection, difference images (e.g. surveillance where changes in a large field of view can be extremely small), optic flow, texture difference images, polygonal approximation of curves and image segmentation. Rosin comments that, for example in the case of edge detection, *"the true edges will just create a flat tail on the non-edge peak in the edge histogram rather than generate a distinct peak of their own"* which is a data description virtually identical to that which our system produces. Additionally the method is shown to work effectively with a range of data sets generated from different sources. These facts, together with analysis of the algorithm revealing it to be impervious to the type of variations that effected the Dunn threshold, make it a good candidate solution.

Rosin's technique, best illustrated visually (see figure 5.8) fits a line to the data running from the top of the noise peak down to the high valued end of the histogram, specifically from the largest bin to the last filled bin in the distribution. To find the threshold point, simple straight line equations are used iteratively to measure the perpendicular distance between the constructed line and the data curve of the histogram. This is done for every x-coordinate (bin), between the start and end of the line, with the longest perpendicular distance being taken to mark the threshold point.

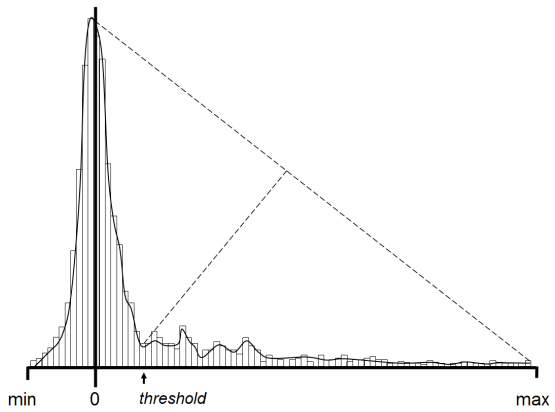


Fig 5.8 – Rosin's Uni modal Threshold Technique. The low boundary value of the marked bin is taken to represent a suitable threshold value.

Fig 5.9 – Operator output (top left, scaled for clarity). Candidate regions resulting from use of Rosin's thresholding technique (bottom left, again scaled for clarity) are found to be in line with a user's perception of the extent of flashlight beams.

Results obtained when Rosin's thresholding technique is applied to the output of the Quotient operator are shown in figure 5.9. As before, with the Dunn and Joseph method (Fig. 5.6), two different flashlights were shone on a highly patterned surface and bounding boxes were drawn round each beam by a human observer. When comparing the results of the two thresholding techniques, the Rosin method is shown to produce regions exhibiting a greater correspondence to those marked by a human observer.

As noted by Rosin, the technique works under the assumptions that the large class (the noise distribution) has lower intensity than the smaller one and that a discernable 'corner' exists at its base, both of which are typically fulfilled in our scenario. Rosin also notes, however, that the method breaks down whenever the fitted line runs too close to the peak, meaning, more specifically, the highest bin in the distribution is very near the peak. This scenario, although happening rarely in our system during active use (flashlights are visible), is in fact certain to occur when no flashlights are switched on and aimed at the monitored surface. In this situation, if the background image is very up to date and no noise occurs, the calculated threshold should flag up no false positives. However, such perfect conditions in practice rarely occur. Instead, it is more likely that the bin containing the brightest pixels (effectively the brightest noise) will be 'labelled' as the end of the second mode of distribution in the histogram which does not, of course, exist. The constructed line, between the distribution peak and this point,

together with the line's longest perpendicular to the data curve, will therefore mark a threshold point somewhere in the middle of the noise distribution. This, in turn, causes the brightest regions (or pixels) containing noise to be considered as candidate flashlights for recognition.

It is important to be able to counteract this weakness in Rosin's method as, ideally; we do not want to present these, falsely identified, regions as candidate flashlights to the recognition process. The specifics of further processing in Enlighten, which exist past the determination of a suitable threshold, are covered in more detail in future sections. In brief however, one such stage involves the fitting of contours around isolated areas found to be above the determined threshold value. At this point, the introduction of a basic filter, applied to such contours, disregards those with either a low circumference or area. This is because such properties are more commonly associated with regions containing noise rather than true flashlights and should therefore not be considered.

For the most part, such minor filtering of results largely counteracts the flagging of false positives in the, formerly problematic, scenario that no flashlights are present. However, in some situations, such as those conditions commonly present in the kiln at Etruria, it was necessary to additionally allow for the optional configuration of user input thresholds. These were not associated with intensity but instead related to geometric properties of the regions returned. For efficiency of testing, such configurations were split into three stages of complexity. In effect, should a candidate region fail on one set of tests, the later, more computationally demanding, checks would not be run. Depending on the requirements configured at runtime, the available tests included placing bounds on the width and height, of a contour's enclosing rectangle, as well as comparisons of these value's ratios to one another. This made it possible to screen out unusually thin candidate regions that are unlikely to be flashlights. Second stage testing similarly applied checks on each region's computed area and the final stage, allowed crude restrictions to be placed on how complex the shape of a contour, surrounding a candidate flashlight region, was allowed to be.

In addition to the above geometric tests, it is also possible to define a restriction based on how closely the intensity data, present in a candidate region, has been matched to previously stored training data. This is described in the next section. It is worth noting that, while it was possible to define restrictions, that enable any combination of, or all, the above described tests to be performed, this was not necessary in the majority of scenarios described in Chapter 4. Ideally it is best practice to use as few, or none of these, restrictions as possible, as this reduces the flexibility of recognition the system can perform. In some cases, such as the Etruria Kiln however, variations in non-uniform global illumination may lead to large areas of background being flagged as candidate regions. This can also be caused by other similar environmental factors, such as shadows or three dimensional objects. Since such cases are effectively a violation of our base assumptions these additional thresholds may be required.

5.5 Training and Recognition

As mentioned in the previous section, the early stages of the recognition technique first involve splitting any presented frame up into a set of candidate regions using Rosin's unimodal thresholding technique. These however are not certain to contain flashlights and some may be disregarded if they are discovered to have eccentric geometric properties. Once this is complete, an active system is initially subjected to a period of interactive training with each flashlight to be recognised. Subsequent to training, comparisons are made between candidate regions and recorded training data, in order to determine a best match to, or discard, the region in question. Recognition is further improved (or regions competing for the same identification dealt with) utilising an algorithm that exploits known motion and position. Each of these stages is now examined in detail.

5.5.1 Normalising Data

Once a suitable threshold has been determined, all data below this threshold is zeroed and the remainder is segmented into non-overlapping rectangular regions which encompass each separate cluster. These are later presented for recognition or as training data for one particular flashlight. To achieve this segmentation, contours, represented by simplified chain codes, are fitted to boundaries of regions above the threshold. These are represented in a hierarchical tree structure so that internal contours, located within others (holes), can easily be identified and removed. Additionally outermost contours are filtered to remove (zeroing the data in those regions) candidates that are likely to be the result of noise, as described in the previous section. This is only done however, if specified in system configuration parameters. Once all erroneous contours have been removed minimal bounding boxes are fitted to those left over and these form candidate regions.

Each of these regions are then analysed as described in section 5.5.2 but before this, the data is normalised using the original stored background image mentioned in section 5.2.2. Such normalisation was not necessary in early experimental versions which did not feature dynamically altered background estimation. However, since dynamic background estimation is required, it is important to account for any changes to a system's operating environment that have occurred since the initial training of recognisable flashlights.

To expand, recall that the operator values we use during the recognition phase are not, as would be preferable, exclusively a representation of the light originating from a flashlight. Instead, these are in fact a representation of that light factored by whatever the ambient illumination is at that point and time. Since our background estimation has now been refreshed to incorporate these changes, any subsequent calculations, using this altered background estimation, will yield different results than those found in our original training set. To counteract this variation, and yield results that are effectively normalized to the conditions of our

original training set, we need to know exactly how much the ambient illumination has changed.

As discussed in Chapter 3, it is impossible to calculate exact ambient illumination values without first having some measurement of the reflectance properties of the surface under scrutiny. It this case however, we in fact only require the *ratio* between the intensity of ambient illumination incident on the scene during training (when our initial background representation was constructed) and that during recognition. Provided we have a measurement for such a value, either globally or calculated on a pixel by pixel basis (to account for potential uneven variations in such illumination), it is possible to normalize a flashlight profile's intensity values. This allows them therefore, to be accurately compared to those acquired during the training phase.

How this is achieved is best demonstrated using the following illustrative example. Here, using values at a single pixel location, we demonstrate and show how we compensate for the effects that variations in ambient illumination have on our results. These values are taken from:

- the originally sampled background estimation e_{bo}
- a foreground image (under identical ambient illumination) containing a flashlight e_{fo}
- a current background estimation (where ambient illumination has changed) e_{bc}
- an associated foreground image, containing the same flashlight at exactly the same position e_{fc}

Usually unknown measurements such as surface reflectance R (unchanging), the actual flashlight intensity I_t and the two levels of ambient illumination (I_{ao} and I_{ac}), are given values in the following example purely for illustrative purposes.

Recall from Chapter 3 that background and foreground image intensity measured at time *x* are given by

$$e_{bx} = R * I_{ax} \tag{5.9}$$

$$e_{fx} = R * (I_{ax} + I_t) \tag{5.10}$$

while flashlight intensity (factored by ambient illumination) is given by

$$I_{tx} = (e_{fx} / e_{bx}) - 1 \tag{5.11}$$

Setting original ambient light (I_{ao}) to 30 and reflectance (R) to ½ in eqn. 5.9 gives

$$e_{bo} = 15 = ½ * 30 \tag{5.12}$$

If illumination of 200 is added by the flashlight (I_t) we see, from eqn. 5.10 that

$$e_{fo} = 115 = ½ * (30 + 200) \tag{5.13}$$

Substitution into eqn. 5.11 gives

$$I_{to} = 6\ ⅔ = 115 / 15 - 1 \tag{5.14}$$

the flashlight intensity (factored by I_{ao}) that is used as training data (I_{to}).

If, over time, ambient light (I_{ao}) increases to 100 (I_{ac}) we have, again from eqn. 5.9

$$e_{bc} = 50 = ½ * 100 \tag{5.15}$$

If illumination of 200 is added again by the same flashlight (I_t), then

$$e_{fc} = 150 = ½ * (100 + 200) \tag{5.16}$$

and flashlight intensity (factored by I_{ac}) measured in the candidate region becomes

$$I_{tc} = 2 = 150 / 50 - 1 \quad (5.17)$$

Without normalisation, recognition is likely to fail as ($I_{to} \neq I_{tc}$). To correct for this, a scaling ratio is calculated that is equivalent to I_{ac} / I_{ao}

$$e_{bc} / e_{bo} = 3 \tfrac{1}{3} = 50 / 15 \quad (5.18)$$

and used to normalize results. Flashlight intensities become comparable

$$2 * 3 \tfrac{1}{3} = 6 \tfrac{2}{3} \quad (5.19)$$

and recognition succeeds.

Implementation of such normalisation does not constitute a large change to the system as processed frames simply have their intensity values adjusted. This is done after the bounding coordinates of candidate regions have been extracted, but before the data in such regions has been processed for comparison with that stored as training data. Normalisation proceeds as follows:

1. Apply Quotient operator to original fore and background images, *AmbOp[e_{bo}, e_{fo}]* storing results as training data
2. Apply Quotient operator to current fore and background images, *AmbOp[e_{bc}, e_{fc}]* storing results as un-normalised data (UN)
3. Apply calculated threshold to un-normalised data to obtain candidate regions
4. Calculate ambient ratio image (ARI) by applying Quotient operator* using original and current background images and globally adding 1, *AmbOp[e_{bo}, e_{bc}] + 1*
5. Factor un-normalised data by ambient ratio image, *ARI * UN = ND*
6. Compare normalised data in candidate regions to training data for recognition

* *Since the ambient ratio image must be calculated every frame, the optimised (see section 5.3) Quotient operator is used for speed due to its similarity to the required calculation (e_{bc} / e_{bo}). The global addition of one to its results merely accounts for the un-required (for this calculation) subtraction the optimised method applies*.

The algorithm employs optimised routines (through use of the Quotient operator) and utilises scaling factors calculated on a pixel by pixel basis. As a result it can deal with non-uniform changes in ambient illumination, is robust and computationally cheap. The method is only disadvantaged by its need to keep a record of the initially constructed background representation; this is not a large overhead.

5.5.2 Feature Extraction, Training and Recognition

Once normalisation has been completed a set of representative features are required which are attained by producing a histogram of Quotient operator values for each region. During training only one region is expected to be visible. Unlike that used for determining a threshold point, region histograms start from a set luminosity (the normalised threshold) and are constructed over only 40 bins. This is because it is not only common that regions contain few pixels, and hence use of a large number of bins would prevent stable patterns from forming, but also to reduce the number of calculations required for recognition during the next stage. Although other potential feature sets could have been used, such as analysis of contour data for example, employment of histograms was considered best due it being a relatively cheap (in processing terms) but robust representation of the intensity distributions making up the beam of each flashlight. During training, for each flashlight class to be used, one hundred samples were gathered at a rate of ten per second (irrespective of frame rate) over a period of ten seconds. For the duration of this period users were encouraged to move the torch randomly across the active surface in order to sample as much of its potential variation against different surface textures as possible. Although tests with larger sample sets have been carried out, any perceived improvements to stability were not enough to justify the increased time required for training. The recognition engine employed requires extra processing to carry out recognition when larger training sets are used.

For every flashlight class, 100 samples with 40 features each are stored together with its class ID ready for use in recognition. After training of all flashlights is complete, the system moves to 'active mode'. In this mode, every candidate region presented is similarly reduced to 40 features and compared to those in the training set. This determines which previously seen flashlight class presents the better match. To perform this comparison, initial work focused on employing the 'K Nearest Neighbours' (KNN) classifier (Fix and Hodges 1951). The classifier, in brief, sums the squared difference of the distance between corresponding features in the unknown sample and each training sample and then orders these summed differences in ascending order (see figure 5.10). Out of these ordered samples, the K best results (in our case a K value of 20 was used, determined empirically) are examined and the class of flashlight found most frequently within this set is taken to be the most likely identification.

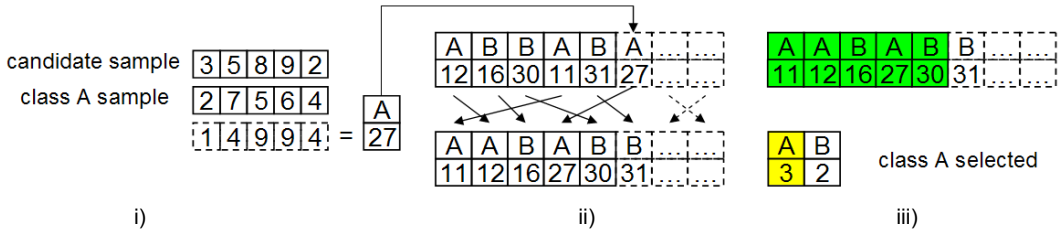


Fig 5.10 – The K nearest neighbours classifier. For illustrative purposes samples only have 5 features and K is set to 5. i) Candidate sample is compared to each stored class sample in turn and squared differences are summed. ii) Results are stored in an array which is then sorted into ascending order. iii) The K best from the sorted array (green) are considered and the most frequently occurring class selected as the most likely correct identification.

Under ideal circumstances, where every presented region in a frame is actually representing a flashlight, use of KNN classifier is found to perform well. However, as has been noted previously, this is not often the case. A presented region set can in practice contain areas of manifest (or actual) non-uniform global illumination variation (not yet accounted for by background adaptation) and/or recently uncovered shadows. Additionally, on semi-reflective surfaces elements of secondary illumination, where the source of the beam (i.e. not the beam itself but the flashlight bulb and reflector) is being partially mirrored (Fig. 5.11), can also appear. Since KNN will always provide a match for each given region, duplicate

classifications will commonly occur in these circumstances – multiple regions may be assigned the same class ID. Some measure of identification confidence therefore, is required in order to help resolve contradictions.

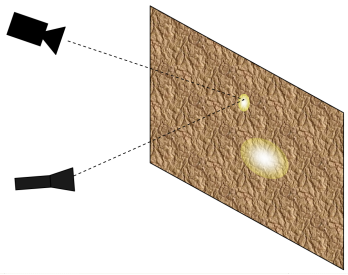


Fig 5.11 – Under certain conditions, a reflection of the bright mirror and bulb of a flashlight is apparent enough, in a camera's view of a partially reflective surface, to be detected as a potential second flashlight.

One way to measure confidence is to present the number of winning class samples found within the K best as a percentage of the total. This would mean that in the case say where the winning class makes up 90% of the samples in this set, its associated region is more likely to be correctly identified, than a region with the same identification whose samples make up only 65%. Such a metric breaks down however when considering that, with a four flashlight system, only six samples of any given class are required to be present in the K set in order to win. If an identically classified region scores over this, say with the minimum winning confidence estimate of just 35%, then effectively we are accepting region identifications which are extremely uncertain. It was also observed that, with just two potential flashlights available as identifiable classes, competing regions often exhibited similar confidence estimates. Under these circumstances, there are in fact only nine possible confidence measurements with which a winning class can be reported. Given this issue, and the fact that use of more than two flashlights can produce low confidence selections, a better metric is required.

In addition, the percentage metric suffers from the fact that even a good confidence measure, for example eighty to ninety percent, bears no relation to how similar a presented sample actually is to those found within the training data. Although the chosen class' samples may represent some of the smallest deviations from the sample to be identified, those deviations may in fact be unreasonably high. If this is the case, the region in question is highly unlikely to contain a

recognisable flashlight beam. A better technique is to incorporate these measured deviations into the confidence estimate. Not only does this eliminate the possibility of two or more regions, that are competing for the same class ID, scoring equally but also allows a confidence threshold to be set. As discussed in section 5.4 such a threshold can be configured based on the observed variation of confidence scores in correctly identified flashlights. When set, any candidate regions found to have distance measures above this threshold are eliminated from consideration regardless of their allocated ID. In these cases, data contained within such regions is grounded to zero, as is also done to data below the global threshold.

There are a number of ways in which the literal difference between training data and candidate regions could be factored into a confidence metric. One possibility is to use the smallest measured deviation from the presented sample, out of those training samples belonging to the winning class (Fig. 5.12i). Alternatively an element of clustering could be incorporated to average the closest matches of the winning class in K, before another class's sample occurs in the sequence (Fig. 5.12 ii). Finally, consideration of the average distance of all the winning class' samples from within K could be used, as this gives an overall impression of how good the match is over a maximum of 20 samples (Fig. 5.12 iii).

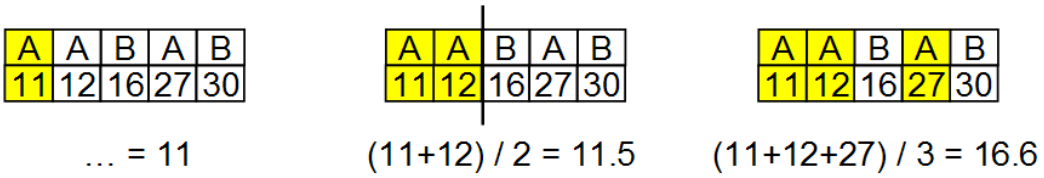


Fig 5.12 – Methods for determining classification confidence metrics with KNN. i) smallest deviation ii) average deviation of first cluster of winning class samples in K iii) average of all winning class samples in K.

The initial technique considered made use of the latter of these three measurements due to it being comparable to the way in which the classic 'Weighted K Nearest Neighbours' (wKNN) classifier is calculated (Dudani 1976, Bailey 1978). In order to factor in the original majority metric, additionally, the calculated average distance was multiplied by the reciprocal of the number of samples associated with the winning class over K. This is simply to ensure a large

matching class presence in the K set gives a lower score as, in the average distance metric, a value as close to zero as possible represents the nearest match. The confidence metric $C_{(wc)}$ simplifies as shown in eqn. 5.20 where $D(x, p_{i(wc)})$ is the summed distance of each sample associated with the winning class $p_{i(wc)}$ found in k, from the candidate region's feature set. Here, the value n represents the number of samples, associated with the winning class, that are found in k.

$$C_{(wc)} = \sum_{i=1}^{k} D(x, p_{i(wc)}) * \frac{k}{n^2} \tag{5.20}$$

This metric in fact holds much in common with the more widely known wKNN technique. Given that the wKNN classifier features our requirement to incorporate each sample's measured distance in its initial calculation, and is more widely known, it would appear the better candidate for the task. Additionally, although a more heavyweight classifier, wKNN is more likely to provide correct initial identifications for candidate regions. This means that, in some cases, resolution between regions which are potentially competing for identical classifications, (see section 5.5.3), may not be required.

$$W(x, p_i) = \frac{\frac{1}{D(x, p_i)}}{\sum_{i=1}^{k} \frac{1}{D(x, p_i)}} \tag{5.21}$$

In wKNN, as with KNN, candidate samples are compared to each class sample in the training set. Feature difference scores are then summed and sorted. The best K matches are then considered in order to perform the final classification. The technique differs however in that the weighted version of the classifier does not take the most frequently occurring class as the winning identification. Instead, each sample has a weight calculated (eqn. 5.21) which is based on how close its features are from the presented candidate sample, in comparison to the closeness of the samples also available within the K set. The winning classification is determined by summing the weights of samples associated with each flashlight

class in the set and the highest scoring of these, ranged from zero and one inclusive, is taken as the most likely identification. The reciprocal of this value can, again, be used as a confidence measure for resolving conflicts as detailed further in section 5.5.3.

Unfortunately, examination of this technique reveals that it exhibits the same issue that is manifest in the use of the KNN algorithm. Although results may present a winning class, that has a weighted majority significantly greater than other classes found in the best K data samples, the average distance of that class from the presented sample may in fact be very large. When only genuine flashlights are presented as candidate regions, such issues can be ignored. Problems arise however, when a region representing a flashlight competes for the same class ID with another region generated by, e.g. a random shaft from a window. It is possible that the spurious region may score more highly than the one containing the genuine flashlight. This is because, although the region genuinely containing a flashlight may match very closely to its own class' training samples, it may also have matched (though not as well) to those of other flashlight classes, hence lowering its score. In the case of the shaft of light from the window, it is equally possible that its presented features matched badly to all the other class' training samples but slightly better to those from the winning class. If comparisons to other classes interfere less, in this region's final score, than they do for the region containing a genuine flashlight, it is likely therefore that the non flashlight region will score higher and hence be mis-classified. In order to counteract this scenario a solution is required which incorporates absolute measurements when calculating each region's confidence metric.

To achieve this, the confidence metric described initially was revisited. That used the average distance of samples in the winning class from the presented sample, factored by the number of times that class appears in the K nearest neighbours. The revised technique again makes use of the average distance of winning class' samples but instead factors this by the score calculated by weighted KNN. Since the new score is not factored by n (the number of samples from the winning class

found in k) the new confidence calculation is given in eqn. 5.22. Here $D(x, p_{i(wc)})$ is, again the summed distance of each sample associated with the winning class $p_{i(wc)}$ found in the K set, from the candidate region's feature set. $W(x, p_{i(wc)})$ is the calculated weight of each sample from the winning class, p_i also found in the K set. In short, the confidence metric is based on the average distance of the chosen class' samples factored by the reciprocal of its weighted KNN score. This not only provides an identification system where confidence measurements are based on physical and statistical factors but also where the smallest values given represent the closest matches. This means they can also have a threshold applied. The algorithm detailing how such confidence measures are used, to resolve identity conflicts, is presented in the next section.

$$C_{(wc)} = \frac{\sum_{i=1}^{k} D(x, p_{i(wc)})}{n} * \frac{1}{\sum_{i=1}^{k} W(x, p_{i(wc)})} \tag{5.22}$$

5.5.3 Temporal Smoothing

At this stage, the system has available a number of regions marking the positions of potential flashlights, each with an initial identification and indication of the confidence with which that identification has been assigned. It is unrealistic, however, to expect a recognition rate of 100%. Our system, in most cases, drives a fully interactive exhibit where the position of each flashlight must control volume and playback of sound samples without latency (see section 5.7). A recognition error prevalent over just a few frames (one tenth of a second) can therefore result in a markedly disrupted user experience. Spurious changes in perceived flashlight identity will cause the media played to change unpredictably, producing a stuttering effect. A way of counteracting, filtering out and allowing for such errors is required.

To enable such filtering to be applied, it is necessary to associate potential flashlight projections between frames – to match each flashlight region extracted from one frame with those detected in the next. Flashlight regions are typically large and so, although they can move quickly and erratically, in most cases the projection of a given flashlight in frame n will overlap the projection of the same flashlight at frame n+1. As flashlights are recognised, independently, in each frame, any filtering performed at this stage can take advantage of both recognition and association results.

If flashlight candidates with the same label are unambiguously associated (i.e. overlap) it is likely, if not certain, that the flashlight in question has been correctly identified. Further processing is required, however, when regions with different labels overlap or if no associated region can be found in frame n+1 for a labelled candidate region in frame n. Both these cases are dealt with using a simple temporal smoothing algorithm over a fixed time window.

Any region which is associated with a region in the next frame, but whose identity appears to change, retains its previous label for a period of N frames. If the new label persists, i.e. is reported by the recognition system for more than N frames, it is assigned to the region at the end of that period. Should the recognition module report the original label within the N frame window, the process resets. The region's identity is again considered to be stable and any subsequent label changes must again be reported for N frames if they are to be accepted. Localised recognition errors are therefore smoothed out. Figure 5.13 illustrates this filter, with N=5. In practise, N=10 has been found to be the most effective value.

The smoothing mechanism also corrects initial recognition errors (Fig. 5.13ii). Flashlights entering the field of view from out of shot are only partially visible for a few frames, and may receive an incorrect label. Once the region is fully visible and has been consistently identified for N frames its new identification will be accepted as true.

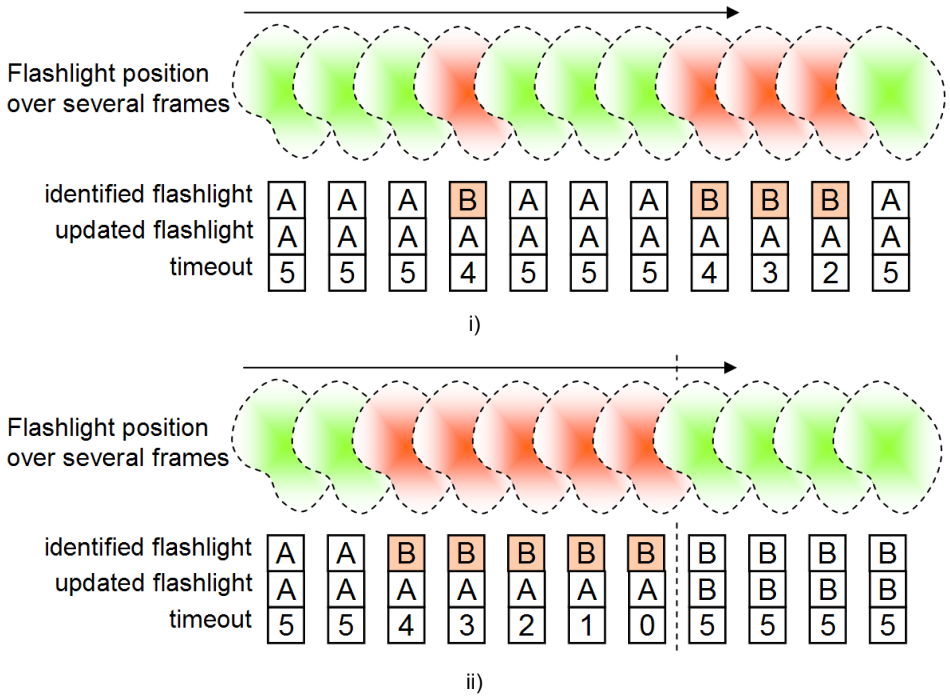


Fig 5.13 – Representations of flashlight position over several frames. Here, red denotes identifications that overlap the last know position of a different flashlight class and green denotes a normal identification. In i), sporadic false identifications of flashlight do not result in a change to its previous classification. In ii), persistent 'false' identifications of a flashlight result in a timeout (5 instead of 10 is used for illustrative purposes) of its previous classification hence a new classification is assumed

Should a labelled candidate not be associated with any region in the next frame (perhaps because the flashlight has been turned off) a similar rule is applied. The data structure generated by that candidate remains in place for N frames, if during that time it becomes associated with another region, processing continues as described above. If not, it is deleted. This mechanism effectively smoothes out the effects of flashlights flickering, perhaps because of poor battery connections, or being momentarily moved off-target, if the user's arm is knocked, for example.

Several motion analysis algorithms (most notably Proesmans et al 1994) ensure consistency in their results by running data association methods in both temporal directions. The equivalent here would be to require each region in frame n+1 to be associated with an unambiguously labelled region in frame n. This, however, places an unreasonable constraint on a flashlight's movement, effectively limiting the speed with which a user can aim a beam to a distance of thirty times its maximum diameter size per second (in accordance with frame rate). While such a

restriction might be acceptable when a system monitors a space where flashlights are aimed at surfaces located in close proximity to the user, it would however cause problems where they are utilised for installations requiring a greater throw. In these situations, since the distance from a beam's origin to its point of intersection with the active surface is increased, the angle of movement required to move a flashlight a set amount resultantly decreases hence the typical speed of a user's handling of a flashlight is likely to be much greater than the proposed restriction would allow.

Despite the proposed technique's ability to deal with false positives while not propagating errors, it has no means to deal with situations in which the same flashlight is (incorrectly) identified at different locations within a single given frame. Each flashlight is represented by a single data structure, if the same flashlight's location is updated more than once, in any given frame only the last valid location for each has any effect on its final recorded position. This can seriously disrupt interaction. The problem was originally observed during the Newark show performance (see Chapter 4) where flashlights, correctly identified as regions that should be triggering targets, were also falsely recognised as being present in the lower part of the image. As regions toward the bottom of a frame are processed last, if their assigned identification conflicts with those regions higher up, the coordinates of the flashlights previously thought to be located in the upper part of the frame are effectively overwritten. This makes it impossible for targets located in the upper part of the active surface to ever actually be triggered. Although one potential approach to counteracting this effect is to simply run sound (or target) control code after each region is identified, this is only partially successful. In this case, a stuttering of playback occurs instead because despite sounds now having the opportunity to commence playback, the program feature that causes them to be silenced after the triggering flashlight has moved away from the sound's associated target, is also triggered within the same frame. This is due of course to the positions of false matches located in the lower half of the active area also running the aforementioned sound control algorithms.

It is clear from the above discussion that, while having a system that is capable of both dealing with temporary fluctuations in flashlight recognition reliability and correcting its own mistakes, it is not practical to allow a scenario whereby the same flashlight can be identified more than once from within a single frame. Instead, the utilised algorithm makes use of the calculated confidence metric for each identified region (discussed in the previous section) applying the following steps in order, to correctly identify each candidate region in a frame without duplicates:

1. For each candidate region, calculate which flashlight is the most likely match and the confidence score of said match using the metric presented in section 7.6.2

2. Where duplications of a flashlight identity occur choose for this identity the best match by selecting from the conflicting regions the one with a confidence metric closest to zero. The regions that are not chosen are saved to a separate list.

3. For each identified flashlight region, if it does not overlap any of the other previously known flashlight positions (excluding its own) update the system's record of this flashlight's position and reset its timeout value. Otherwise add it to the list of unidentified regions.

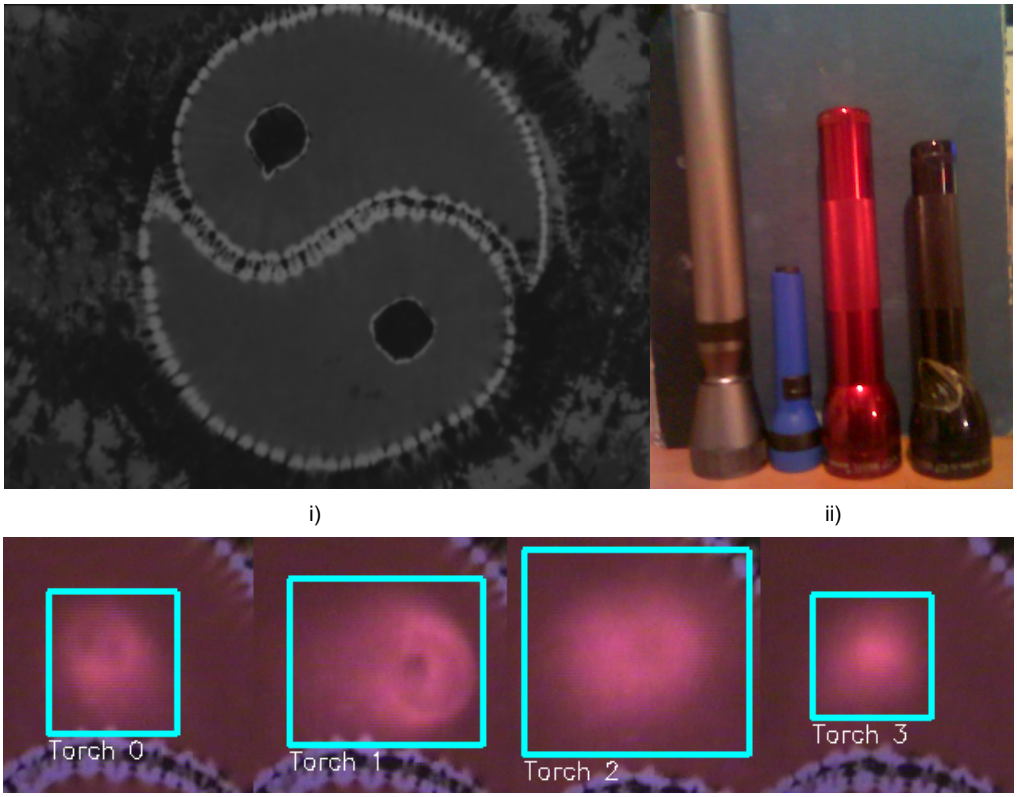
4. For each the remaining unidentified regions, if it overlaps a previously known flashlight position that has not already been updated, update the system's record of that flashlight's position with the overlapping region's position but **do not** reset its timeout value.

Application of the above logic not only aids with the identification of candidate flashlight regions but also allows us to make use of any initially unidentified regions while at the same time maintaining the original technique's strengths of counteracting erroneous updates yet not allowing identifications made in error to ever propagate for longer than a ten frame period.

5.6 Recognition Performance

To assess the recognition performance of Enlighten the following experiment was performed within the MRL, University of Nottingham. A camera was aimed at a heavily textured surface and Enlighten trained to recognize 3 sets of flashlights,

containing 2, 3 and 4 distinct flashlights respectively. Figure 5.14i shows the background texture employed throughout. Figure 5.14ii shows the physical flashlights used and figure 5.14iii shows the appearance of each of the flashlights when projected onto a reasonably uniform section of this background.



i) ii)

iii)

Fig – 5.14 during evaluation of recognition performance, background texture i) was used with flashlights 3-0 (as shown left to right) ii) which cast profiles 0-3 iii)

Following each of the three training sessions each (trained) flashlight was played, one at a time, over the test surface for a period of 15 seconds. In each case video was captured at 20 fps and the resulting 300 frames run through Enlighten. The recognition methods described above were applied to each frame, resulting in a total of 2700 recognition attempts. To allow the effect of temporal smoothing to be assessed, flashlight ID was recorded both before and after temporal smoothing was applied. Confusion matrices for each of the six cases (2, 3 and 4 flashlights, with and without temporal smoothing) are shown below.

Tables 5.1-5.3 show recognition results obtained before temporal smoothing. Recognition rates are consistently very high. There is some degradation in performance as a larger set of flashlights is considered. This is to be expected and reflects the increased likelihood of members of a larger training set appearing similar. Though these results are impressive, it must be remembered that flashlight recognition is applied independently to each frame of a real time video stream. At 20fps a recognition rate of 99.9% would result in a recognition failure every minute. Smooth interaction is unlikely to be possible in these circumstances.

		Actual Flashlight ID	
		0	1
Reported Flashlight ID	0	99.33%	1.66%
	1	0.67%	98.34%

Table 5.1 – Flashlight recognition results obtained when two flashlights were trained. Reported flashlight IDs were gathered before temporal smoothing

		Actual Flashlight ID		
		0	1	2
Reported Flashlight ID	0	99.69%	1.88%	
	1	0.31%	76.1%	17.76%
	2		22.01%	81.93%

Table 5.2 – Flashlight recognition results obtained when three flashlights were trained. Reported flashlight IDs were gathered before temporal smoothing

		Actual Flashlight ID			
		0	1	2	3
Reported Flashlight ID	0	100%	0.58%	6.35%	
	1		81.55%	29.29%	
	2		17.87	64.16%	10.56%
	3				89.44%

Table 5.3 – Flashlight recognition results obtained when four flashlights were trained. Reported flashlight IDs were gathered before temporal smoothing

Tables 5.4-5.6 show recognition results obtained after temporal smoothing. When sets of two and three flashlights are involved, recognition rates reach 100%, suggesting that previous recognition failures were only intermittent. Some failures occur when four different flashlights are employed. Again this reflects the similarity of the flashlights employed. Though this situation could be improved by the addition of further information to the flashlight description, in practice interactive installations rarely require more than two or three flashlights to be distinguished.

		Actual Flashlight ID	
		0	1
Reported Flashlight ID	0	100%	
	1		100%

Table 5.4 – Flashlight recognition results obtained when two flashlights were trained. Reported flashlight IDs were gathered after temporal smoothing

		Actual Flashlight ID		
		0	1	2
Reported Flashlight ID	0	100%		
	1		100%	
	2			100%

Table 5.5 – Flashlight recognition results obtained when three flashlights were trained. Reported flashlight IDs were gathered after temporal smoothing

		Actual Flashlight ID			
		0	1	2	3
Reported Flashlight ID	0	!00%		6.17%	
	1		100%		
	2			93.84%	
	3				100%

Table 5.6 – Flashlight recognition results obtained when four flashlights were trained. Reported flashlight IDs were gathered after temporal smoothing

5.7 Triggering Logic

In the majority of the installation scenarios described in Chapter 4 the final stage in the process is the triggering and control of audio playback. Following on from the observations made during the cave experiments (Chapter 2), it is clear that in the majority of cases simply triggering a sound is not enough to allow a user to associate it with its correct location. In the case of Etruria's so called 'hidden targets' for example, the typical search method (of 'painting' a surface with a flashlight) would often leave a user misled as to the true location of the target they have found. This was due either to fast flashlight movement, delay in the sound clip reaching a noticeable volume (i.e. it is not immediately apparent that a sound is playing) or the varied reaction times of the user. The obvious solution to this issue was to simply cease audio playback should a flashlight beam leave a target region, however initial experimentation with this technique proved problematic.

One of the largest contributions to the usability difficulties caused by the implementation of this technique was simply that of sounds failing to reach audible levels before being silenced as the flashlight leaves the target. One solution could of course be found in the issuing of guidelines to audio creators (Etruria, Intech and MAGNA were all responsible for making their own content) to ensure ample volume at the very start of each clip. In many cases, this allowed users to hear the audio clips, even when flashlights were being moved at quite high speeds. Audio was often, however, heard only momentarily, and it was observed that users would therefore still exhibit difficulties in relocating the exact point that had triggered it.

In addition to these difficulties, as audio was either completely on or completely off during playback, users had no indication of how close they were aiming their flashlight to the edge of a sound triggering region. This, especially for young children struggling to control large and relatively heavy flashlights, could often be a source of frustration. The resultant wobble (see Chapter 2) if occurring near a triggering boundary would frequently cause an audio clip to restart. This made it hard for younger users to experience longer clips in their entirety, for example as commonly found in the Explore the Universe exhibit at the Intech Science Centre.

The revised solution employs the use of volume levels as an indication to the user of how far away they are from a content hotspot thus indicating the danger that excess movement, when audio has become very quiet, may potentially cause a clip to restart. Additionally, the use of volume to represent proximity serves as a guide to aid a user in locating the exact source of a triggered clip. This means that it is always very clear exactly which content is associated with which location or point of interest.

Algorithmically, the metric used for determining volume level is based on percentage overlap between the region encompassing a flashlight beam and that used to represent a target area. Alternatively, proximity between the centres of these regions might have been used. This was rejected, however, as such a

measurement allows for only one, very exact, location where a clip can be experienced at its maximum level. Instead, use of region overlaps has the benefit of being independent of either flashlight or target size as the largest of the two values is always chosen and mapped directly to a percentage of the maximum volume level.

To trigger a target to start playing, fifty percent of either the target or flashlight must be overlapped and this must be sustained for three frames if the flashlight has only just been ignited. This delay is required in order to account for the observed situation that it takes more than the time space of a single frame for a flashlight to become fully lit; therefore it is necessary to delay attempted recognition until it is certain that this state has been achieved. Once initiated, audio clips will continue to play at a volume determined by the activating flashlight's position. In order to counteract "wobble" each audio clip is also granted a ten frame timeout or grace period similar to that used for flashlight recognition. Such a feature ensures that should a user experience a sudden but brief jolt to their aim, assuming they can reposition their flashlight beam again quickly, audio playback, although momentarily silenced, should continue from its previous playback position.

Aside from MAGNA, whose requirement for integration with their existing infrastructure meant that the software functioned similarly to that used within the Nottingham Caves, the above logic for controlling user interaction was found to be extremely beneficial in all installations. Intech in particular, with its almost exclusive use of hidden or discoverable targets, additionally utilised a cross fade mechanism between featured content and an ambient background sound. This, being utilised in an environment where several installation are positioned in close proximity to one another, helped to give clear indication to the user as to when they were searching within their correct content area.

5.8 Summary

The interactive flashlights system Enlighten has been deployed in the various installations discussed in the previous chapter. We have described the architecture of Enlighten and looked in detail at each of its components, focussing on the issues raised by the various installations and the design decisions taken as a result. In the next chapter attention is given to the remaining issues that have been noted in Chapter 4 and the technical discussion above, providing full explanation for the cause of these problems and exploring some potential solutions to these occasionally arising situations.

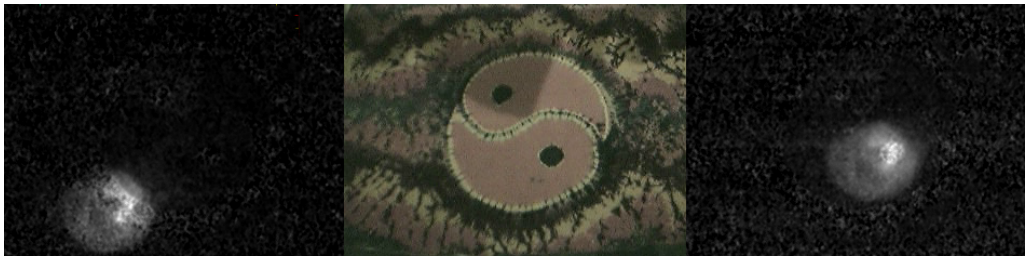
Chapter 6

Non-Uniform Illumination and Dynamic Range

The previous chapter presented an overview and in depth examination of the various components and techniques used within the interactive flashlights system. These were designed to meet the varying requirements of the experimental installations discussed in Chapter 4. A recurring problem was noted during the descriptions of several of the techniques used; in several cases the output of the Quotient operator was clearly not independent of the reflectance patterns present on the target surface. Despite such problems often being overcome by, for example, use of normalisation procedures and rolling background estimation it is important to understand why they occur and ask if there are any additional solutions that might be employed in order to prevent them. During initial experimentation, it was noted that such situations most commonly occur when the active surface exhibits extreme variations in background reflectance. We will now examine how this and other environmental factors affect system operation, and explore potential solutions.

6.1 Surface Reflectance and Dynamic Range

Figure 6.1 illustrates a problem found in each of the experimental environments described in Chapter 4, most commonly when the background surface is marked by large changes in surface reflectance. Figure 6.1ii shows the heavily textured surface used in Chapter 5, while figures 6.1i and 6.1iii show examples of the operator output obtained as the flashlight beam is moved over the surface. Note that, though the operator highlights the flashlight beam, in each case the operator response is not entirely independent of the underlying surface texture.



i) ii) iii)

Fig 6.1 – i) and ii) Scaled responses of the Quotient Operator as the same flashlight moves about a heavily textured surface. In iii) the black circle (centre ii) is clearly visible in the operator response, as is the dark band (lower centre left ii) in i)

This is problematic. If operator response cannot be relied upon to be independent of surface reflectance, flashlight recognition also becomes unreliable. The effect is also unexpected; however there are a number of ways in which it could arise.

An effect similar to that seen in Figure 6.1 could arise from camera shake. If the camera moved between acquisition of the fore and background image the operator would be applied to pixels sampling different points on the target surface, with potentially different reflectance values. Camera motion, however, would produce an effect that emphasised reflectance changes. Pixels would experience the largest change in the reflectance values they sampled when camera motion moves their field of view from one side of a reflectance edge to the other. As figure 6.1 shows, this is not the effect produced. Camera movement is not the cause.

Another possible cause is a change in ambient illumination. The likelihood of a pattern of ambient illumination matching the reflectance features of the target surface is, however, clearly very close to zero. It is also logically possible that reflectance has changed between capture of the fore and background images. This is clearly impossible in the real world.

The only other explanation is that, although surface reflectance hasn't changed it instead only *appears* to have done so. According to the Quotient Operator's equation (see Chapter 3 eqn. 3.20), assuming a reliable measurement of the fore and background intensity is made at every pixel, the reflectance factor should be the same in each, and so cancel out. Given the correlation between reflectance

patterns and operator response, however, we must assume that, under certain circumstances, this is not actually happening.

The only logical way in which the reflectance could have an effect on the operator output therefore is if, somehow, it is measured differently in the fore and background images. Reflectance independent of other factors isn't actually quantified of course, as the only measurements our cameras provide are pixel intensities. Variations in the way in which intensity data is captured between the fore and background image will, however, affect the operator output as changes in illumination and/or reflectance would. The remaining question is in what way has the image acquisition process changed.

Having established that it is indeed possible for variations in background reflectance to have an effect on operator output, the question remains as to why these distortions in intensity measurement occur and if there is any way of counteracting the problem. Examination of operating environments and utilised equipment reveals the most likely cause for such sampling inaccuracies. This is the fact that we are attempting to capture scenes which exhibit a very high dynamic range, using standard cameras capable of sampling over only a limited range of light levels.

High dynamic range images arise frequently in flashlight installations. Many are semi-dark, having only low levels of ambient illumination and containing at least some dark surfaces. Images of target surfaces captured under only ambient illumination (i.e. background images) are therefore typically obtained by sampling only low intensity light. In contrast, flashlights can be very bright, raising the level of light reaching the camera considerably. If the interactive targets used are intended to be visible to the user, it is tempting to make them noticeably brighter or darker than the background. Light targets on dark backgrounds have been particularly common. The level of light transmitted to the camera when a bright torch is played over a light target (in a foreground image) can be very much higher than that generated by ambient light leaving a dark background (in a

background image). If the targets used have a specular component the range of light levels produced is increased even further.

Although domestic webcams and other commercial (e.g. CCTV) cameras now provide high levels of performance at comparatively low cost, each can sample only a limited range of light levels. Camera parameters (aperture, shutter speed, etc) allow the set of light levels being sampled to be controlled to some extent. It is often the case, however, that the range of illumination levels that can be sampled by a given camera is closely matched to or even smaller than that which might be generated from a particular flashlight installation.

It is important when setting up a flashlight system that the camera parameters be matched to the dynamic range of the installation. A number of problems can arise.

It has already been identified through previous experimentation that allowing flashlights regions to become saturated inhibits recognition due to each incident beam having no identifiable features, other than being made up of a white uniform region. To counteract this, it was necessary to implement an additional stage in system configuration whereby camera sensitivity levels are adjusted (by altering 'shutter speed' etc.), against scenes where flashlights are present, until the amount of light being captured is sufficiently low that saturation does not occur.

Configuring the camera simply to avoid saturation, however, is problematic, as was discovered during the Newark show (see Chapter 4). Such adjustments can cause frames, where flashlights are not present, to be so dark that they capture little information about of the scene under scrutiny. Camera parameters chosen to prevent flashlights saturating when shone on reflective targets can result in intensity values at or near zero from large regions of the background image.

One solution is to utilise lower intensity flashlights so that a brighter background representation might be captured without saturating the foreground images. However, in situations where illumination changes can be sizeable, problematic

conditions can also occur over time. Camera parameters and flashlight intensity must take into account any likely changes. Predictive configuration can be used. Initial setups undertaken during cloudy or dark environmental conditions, for example, would utilise only a lower portion of a camera's full dynamic range with the upper portion only being used if conditions become bright. When setting up in bright conditions, the reverse is true.

To select flashlights and set camera parameters for a given flashlight installation is to attempt to find a trade-off between the factors discussed above. The camera and flashlight combination must allow the patterns of light in the flashlight projection to be reflected in the operator output, but also allow background images to adequately represent surface markings and texture on the target surface. This is difficult to achieve, and often results in the entire dynamic range of the camera (i.e. the lowest and highest recognizable intensity values) being used.

Given a camera with a truly linear response to input light levels, utilization of extremes of camera range, while hard to set up, should not in fact cause problems. It is likely however that the intensity response of the camera used may not be linear over its entire range. This is illustrated in figure 6.2 which shows that, while linear over the majority of input light levels, response may vary toward the extremely high or extremely low ends of the scale.

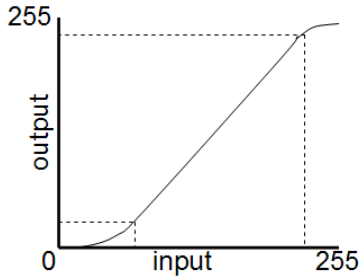


Fig 6.2 – Representation of a potentially semi-linear camera response. Input illumination (arbitrary scale) as compared to their equivalent reported (output) intensity values from the camera

Non-linear camera response can account for the effects seen in figure 6.1. If the fore and background images are both acquired by a camera whose response is linear over the light intensities involved, taking the quotient of corresponding

pixels will produce a response in which reflectance is cancelled out. If, however, either or both is produced by sampling the non-linear part of the camera response, this need not be the case.

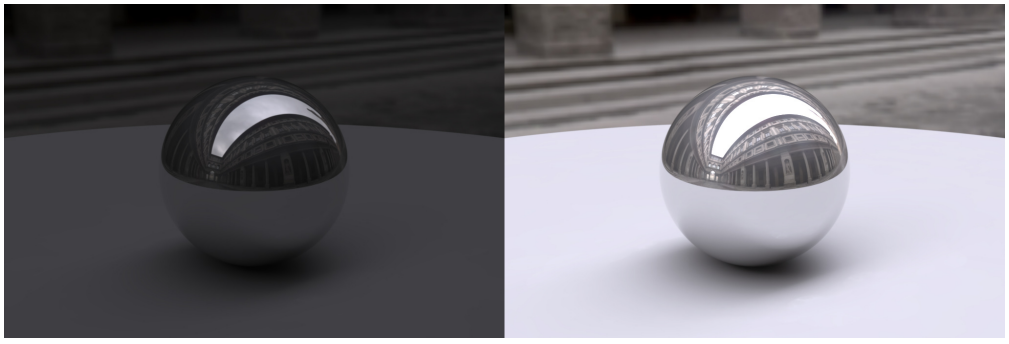
Consider a situation in which the background image is reasonably bright, and the pixels it contains arise from the linear segment of figure 6.2 , but foreground images arise which sample the extremely bright, non-linear section of the camera's range. In effect the camera has applied two different scalings to the combined illumination and reflectance values making up the light emitted from the target surface. The quotient of fore and background pixels will contain a component in both illumination and reflectance that is proportional to the difference between these scale factors. In this example, the effect on the interactive flashlights system may be comparatively minor. Only the brightest pixels of the brightest flashlight will be affected, and then only when that flashlight is directed towards a more reflective target. There may be some distortion of the operator output, which could disrupt recognition, but the pattern recognition engine may be able to deal with this.

Potentially much more disruptive are situations in which the brighter regions of the foreground images are obtained from the linear part of the camera response, but the background image is sampled from the lower, non-linear section. In this case, a different scaling will be applied to pixels gathered across a much larger proportion of the image plane. Most, if not all, of the background image pixels will effectively be obtained under different imaging conditions to most, if not all, of the pixels in flashlight regions. The operator will therefore produce responses which incorporate a component dependent on surface reflectance across the majority of each flashlight region, causing severe disruption.

Given our issues of insufficient dynamic range, the most obvious solution is to make use of specialist high dynamic range cameras. These produce floating point intensity values by combining multiple images acquired using different shutter speeds (Amtel Corp).

Traditionally, monochromatic images consist of pixels represented discretely by one of 256 possible intensity levels. Under normal circumstances such resolution is acceptable. However in our case, because it is important to gain representations of both extremely bright and extremely dark objects, we are often forced to use background images which comprise just a few intensity levels at the low end of the range. These are required to accurately represent a wide variety of textures and markings on the target surface, which can result in information being lost.

Specialist high dynamic range cameras combine several images taken from the same point under different exposures (see figure 6.3). Taken together, the resulting image set represents intensity with a much higher resolution that is achievable given only a single standard image. The image set can be combined to produce a single, floating point image (Debevec and Mallick 1997). Unfortunately, fast response requirements and the environmental conditions under which the system is commonly deployed make use of such cameras impractical to this application. Not only is such equipment prohibitively expensive, but multi-exposure cameras do not operate as fast as traditional ones and also are not as durable or robust.



i) ii)

Fig 6.3 – Images (*source: wikipedia*) taken under different exposure settings. In i), it is possible to make out individual clouds in the sky which appear as a solid white region in ii) Conversely in ii) details beyond the pillars are apparent but appear black in i).

One alternative to the use of high dynamic range cameras is to either acquire (e.g. from the manufacturer) or construct (through some form of calibration process) an accurate representation of the response curve of a standard camera. This would

allow some correction of image intensity values to take the non-linear sections of the response curve into account. Accurate models of camera response are, however, extremely hard to obtain. Furthermore if, as is often the case, the non-linear part of the curve is flat, or has a low gradient, information will be irretrievably lost. The mapping between light level and reported intensity will be many to one.

6.2 Spatial Variations in Initial Ambient Illumination

High dynamic range is one of two environmental factors which can disrupt the operation of the interactive flashlight system. Ambient illumination that varies significantly across the field of view at system setup is the second.

The normalisation process described in Chapter 5 can compensate for changes in ambient illumination over time. Normalisation scales the operator output to produce values comparable with those obtained during training. If, however, ambient illumination is not spatially uniform when training is carried out, problems arise.

Though the operator developed in Chapter 3 is independent of surface reflectance, its output at a given pixel is scaled by the local ambient illumination value. As the actual contribution of ambient illumination is unknown, it has been assumed throughout the work reported here to be a constant. If ambient illumination is evenly distributed across the field of view when the initial background image is acquired, this is acceptable. If, however, the system is set up in an environment in which ambient illumination is not spatially uniform, the operator response at different points in the image will vary to reflect this. As a result, any flashlight profiles, which are located inline with such variations, will be arbitrarily scaled by the different values of ambient illumination that are present there. Recognition may be affected as a result.

Under most circumstances such as plain wall displays and the like lit by diffuse illumination, the distribution of light hitting a scene should be approximately uniform over small regions and if it is not, those changes that do occur should only have negligible effect on the flashlight profiles obtained (Chapter 3). Situations exist however where it is impossible for scene illumination to be spatially constant. For example, in the presence of either well defined shadows or when a camera's field of view features large areas containing very varied lighting. Given these circumstances, although steps can be taken to avoid such problems, either by using a region mask (Chapter 5) or accepting limitations on the flashlight system's durability, it is of interest to explore how the initial distribution of ambient light over a scene might be determined. This information could then be used, should the spatial variations in ambient light become considerable enough to have a significant effect on system stability, to normalise operator output.

As is well understood, it is impossible to actually model the exact level of light hitting a surface without good knowledge of the underlying surface reflectance properties (Ullman 1975). Techniques such as the Lightness computation (Land 1971, Horn 1974, Blake 1985) might be used to approximate such a description but not only does this assume locally constant surface reflectance (something we cannot), but also can only provide descriptions of surface reflectance as ratios of which ever region is the most reflective. Given that illumination ratios are all we need to compensate for changes over time however (Chapter 5), if a Land and Horn style model of spatial illumination rather than reflectance ratios could be produced, this would provide a pixel by pixel illumination map similar to the Ambient Ratio Image used in Normalisation (Chapter 5).

The Quotient operator may be used to estimate the relative ambient illumination between two points in the field of view. We first acquire a background image as before, in which the target surface is only illuminated by (possibly non-uniform) ambient illumination. An additional light source is then activated, which casts an even amount of illumination across the field of view. Applying the operator to

these two images gives an output at each pixel which is proportional to the ambient illumination at that pixel scaled by the light provided by the additional (flat) source. Taking the ratio of the operator outputs obtained at two image locations cancels out the component in the intensity of the light source, leaving the ratio of the original ambient illumination between those two points.

Some method of integrating ambient illumination ratios across the image (Land 1971, Horn 1974) would be required, however, and there are clear practical problems associated with reliably adding a uniform amount of illumination. One possible approach would be to replace the single uniform light source with a moving flashlight, tracked over time. If a starting point could be identified that is uniformly illuminated, and tracking could be achieved to pixel accuracy, responses from corresponding (i.e. identically illuminated) points in two flashlight regions could be combined as described above to estimate an ambient illumination ratio (Figure 6.4).

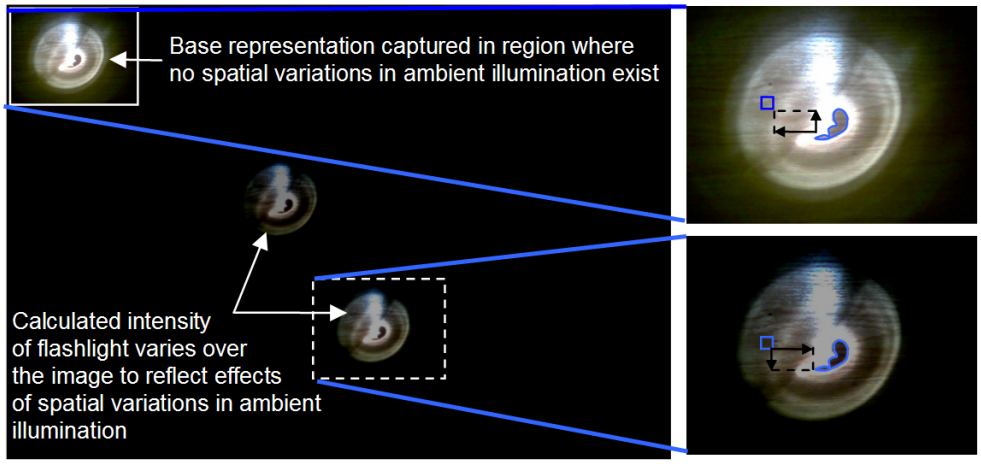


Fig 6.4 – During a sweep, for each calculated intensity point in a located flashlight profile, its equivalent must be found in the profile's base representation. This allows the ratio between the measurements of these two, known to be equal, intensities to be determined. As shown, equivalent points between profiles could be determined by measuring their distances from features that are recognisable even when measured intensities vary. In the example given, the lower corner of the irregularly shaped dark centre of the flashlight beam is used as a reference point to measure to and from.

Despite the fact that spatially non-uniform global illumination levels might have an effect upon operator results, given the non-perfect profiles gained from experiments discussed in Chapter 3, it is more likely that the largest detriment to

system performance, experienced commonly, is instead due to issues discussed in section 6.1. Methods of compensating for variations in ambient illumination during setup will not be considered further at present.

6.3 Conclusion

Two environmental factors have been identified which can adversely affect the performance of the proposed interactive flashlight system. Operator responses that are strongly correlated with surface reflectance are almost certainly caused by quantisation effected by the non-linear response of the utilised cameras, when exposed to scenes of low intensity. In addition to exploring potential solutions to these issues the question of how to account for the possibility of non-linear spatial illuminations was also investigated. However none of the proposed methods, that might allow compensation for such detrimental effects, have been found to be practically applicable. This effectively leaves a system that is liable to decreased levels of performance or stability under certain conditions. With a better understanding of the nature of these degradations, however, steps can be taken in order to minimise or avoid their occurrence.

6.4 Summary

In this chapter we have investigated the exact nature of the technical issues that were raised but not examined in detail in the previous Chapter. To this end, likely causes for such problems have been identified, possible solutions discussed and conclusions drawn as to what might be done either to compensate for, or prevent such issues from occurring. In the next chapter we examine potential future enhancements to the presented system, new directions and applications to which it might be adapted and present final conclusions regarding this work.

Chapter 7

Conclusions and Future Work

In this thesis we have examined the feasibility, development and deployment of visually tracked flashlights as interaction devices. To this end, we have first considered similar or related work with regard to vision based interfaces (Chapter 1) and gone on to report on early trials with three prototype interfaces of this kind (Chapter 2). Analysis of each of these trials showed flashlight interfaces to be feasible (both technically and from a usability point of view) and that participants found them attractive and enjoyable to use. Attention then focussed on the stable description of the patterns of light projected by flashlights. Several methods were developed and evaluated and the Quotient operator (Chapter 3) selected for further study. The deployment of a new flashlight interface in a number of different experimental environments is then discussed, highlighting each location's technical issues and challenges (Chapter 4). The techniques used to develop a suitable interface for use in these situations are then examined in detail (Chapter 5). Finally, irresolvable issues are presented, with consideration of their effects and potential methods by which they might be solved (Chapter 6)

We shall now examine possible future developments of the current interactive flashlight system, preliminary work towards some of which has already been undertaken. Since there are a number of directions which such developments could take, these shall be grouped into the broader categories of:

- Technical developments, relating both to improvements to the existing system and potential functional extensions
- Use of flashlight interfaces in new, physically different, environments
- New applications for the flashlight-based interactive technology

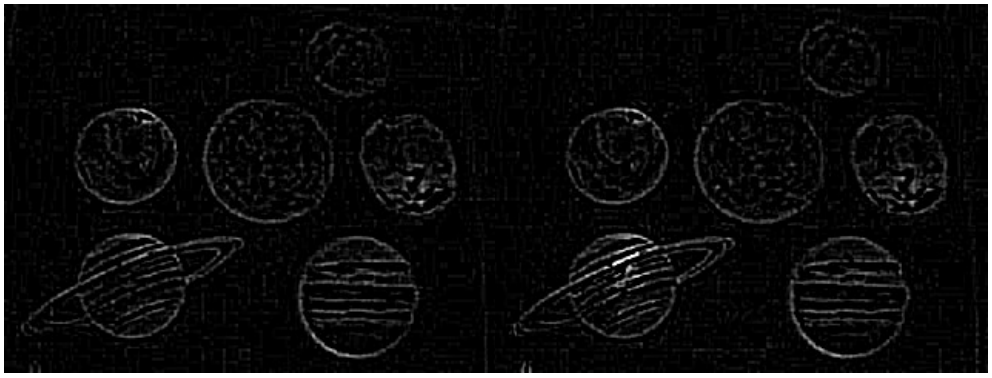
7.1 Technical Developments

7.1.1 Improvements to the Existing System

One of the most obvious areas in which Enlighten might be improved is in the techniques used for flashlight beam recognition. There have been significant developments in object recognition in recent years, most notably the development of boosted classifiers by, for example, Viola and Jones (2001). Though the current k-nearest neighbour method has been found to be effective, further work towards determining the applicability of more recent techniques to the task of flashlight recognition would be of benefit. Alternatively, other avenues by which recognition might be improved lie in the use of principal component analysis (PCA) based methods such as Cootes and Taylor's Active Shape and Appearance Models (1995). Such models might produce a more appropriate space in which to compare current (input) data with that found within training sets.

Additionally, PCA could also be used to provide a more detailed analysis of the training data acquired during a system's initial configuration. This might be utilised in order to guide Enlighten users as regards the suitability of candidate flashlights to be used with the interface. To expand, PCA would fit lower-dimensional coordinate frames to training sets that contained high dimensional data points. Such coordinate frames are represented as sets of eigenvalues, and eigenvectors of matrices of covariance values. Eigenvectors provide the axes of the lower dimensional space, eigenvalues measure how much data variation each eigenvector accounts for. By comparing the eigenvectors and values produced by different flashlights' data sets, it may be possible to automatically identify those which appear too similar to be reliably recognised. These should therefore not be used in an Enlighten installation. If such automatic identification is not possible, alternatively, the standard method of displaying the modes of variation of the data sets might instead be exploited. This could be used to produce a tool which allows installation designers to explore data sets, and decide for themselves, if two flashlights are unacceptably similar.

Related to recognition, in addition to the currently used technique for discarding 'non-flashlight like' regions, based on their similarity (or lack of) to known flashlights, a further method might make use of the inherent properties of flashlights for such classification. There is a fundamental difference between the appearance of flashlight projections in an image and the appearance of an occluding object. This is that an (opaque) occluding object will commonly disrupt the pattern of edges that exist in the background image but a flashlight projection, although having some effect on the pattern, should not disrupt it unduly. A region classifier, based on the comparison of fore and background image edge maps, may therefore prove feasible. Although initial experimentation with the concept has shown potential (Fig. 7.1), further development of the technique is required.



i) ii)

Fig 7.1 – Edge response patterns recovered from both fore ii) and background i) images used in Enlighten. In the foreground image, a flashlight beam is positioned over Saturn (bottom left). Comparison of the two images shows the presence of the flashlight has only a minor effect on the strength of the edges recovered beneath it and does not disrupt the pattern itself, as an occlusion might. This might be used to discriminate between image regions containing flashlights and those containing occlusions

7.1.2 Extensions to the Existing System

One of the most desirable extensions to the current flashlight interface presented by Enlighten, would be to detect and exploit the rotational orientation of flashlight projections. This appears possible because projections of flashlight beams onto target surfaces are usually asymmetric, either due to their profile patterns being affected by imperfect reflectors/filaments, or because the flashlight is held at an

angle, hence generating an elliptical projection. As shown in figure 7.2, some initial work has already been carried out towards this goal, by applying PCA to flashlight regions identified in operator output. For each region, by selecting only the eigenvector associated with the largest eigenvalue, it is possible to identify the major axis of a projection pattern. The orientation of this can be used to represent the orientation of the flashlight itself, which in turn might be used as an additional interface control in future applications.

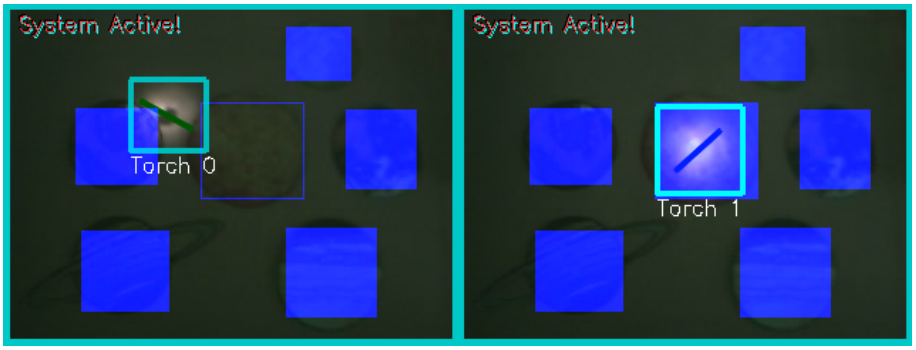


Fig 7.2 – Experimental trials with use of PCA to recover a flashlight beam's orientation. Lines through the centre of beams marked the major axis of symmetry in their projected pattern.

The trial implementation of this functionality was carried out independently on individual frames with no knowledge of previous frames' results being utilised. Although the technique showed potential when used with flashlights projected onto a relatively plain background, the manifestation of reflectance features in operator output, that sometimes occurs in highly patterned areas (Chapter 6), occasionally lead to erratic results. Development of the technique, using knowledge of results over time or filtering for only the strongest responses may, however, lead to a deployable method.

It was noted during early experimentation that not all flashlights used were asymmetric enough for this technique to work reliably. The method may still have worth, however, for collaborative use of flashlights. In this case, two flashlight beams located close to one another (so as to overlap) would give a strong response allowing angles to be inferred from the rotation of one flashlight around the periphery of another. The location where this occurs might be used to dictate the meaning of the gesture, for example, manipulation of symbolic control dials.

Regarding gestures, as noted in Chapter 2, there is evidence (from the trials carried out) to suggest that the ability to recognise movement related gestures (in particular those that are carried out naturally without instruction), might have been beneficial to the Sandpit installation. For this specific installation, detection of the naturally occurring 'moving flashlights in tight circles' gesture, might have been exploited to accelerate digging. When regarding such gestures, flashlights are effectively reduced in complexity to mere pointing devices. Exploration of the techniques used to detect mouse gestures therefore (Pirhonen 2002, Opera Software – example gestures), may prove to be similarly applicable to detecting those made using flashlights.

7.2 New Environments

7.2.1 Projections

During the experiments with the StoryTent and Sandpit, flashlights were used to good effect with projection screens. These systems employed a simple and fairly limited detection mechanism however, through which flashlight recognition was not possible. Application of Enlighten in its current form to such environments is similarly not possible. This is due to its requirement that backgrounds remain (relatively) static.

Extension of Enlighten to allow direct interaction with projections however, might be possible if the projections themselves are constrained. In the case of slide projection for example, *n* background images might be captured during setup and each of these could be dynamically used when its corresponding slide is displayed. Since the system itself would be used to control movement between slides, it would also know to use a new background image, whenever a slide is changed.

7.2.2 Virtual Worlds

Another environment where flashlights might find application, is in their use in illuminating virtual worlds. Essentially, shining a real flashlight into a virtual world, to illuminate it in the exact same way the flashlight would illuminate a scene in reality. Specifically the same pattern of light would be apparent. Early work towards this goal, carried out by Helen King (2006), employed the Quotient Operator to recover flashlight illumination incident on a flat surface. This surface was considered synonymous with a transparent one located within a virtual world, through which illumination, based on that from the flashlight, was emitted. To position the virtual light source correctly, the technique used contour analysis of the flashlight's beam, to infer its relative location and angle in the real world. (Fig. 7.3)

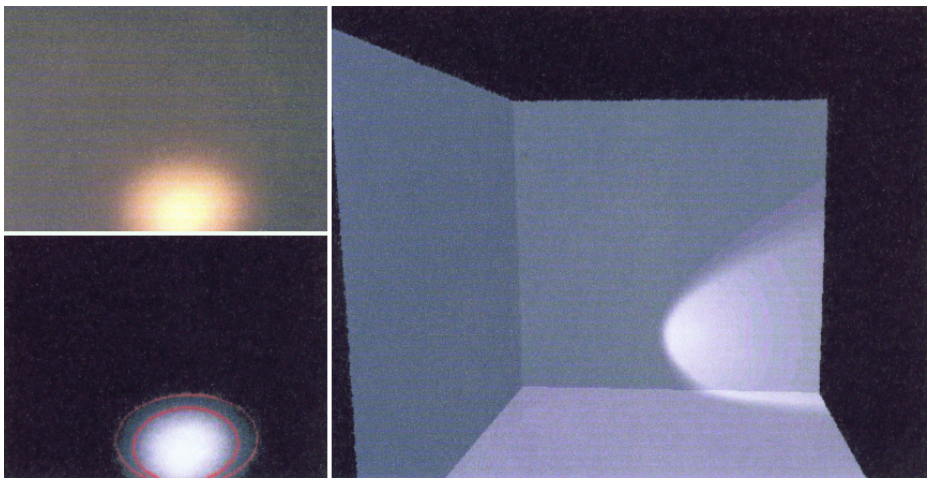


Fig 7.3 – Early work towards shining real flashlight beams into virtual worlds: Top Left: original flashlight profile, Bottom Left: fitting contours to smoothed operator output to recover flashlight position and orientation, Right: the effect of the flashlight in a virtual world

Although demonstration of the method indicated technical feasibility, 3D recovery of the flashlight's position was problematic and this led to the virtual light source moving erratically. Further work is required therefore, to see if such position data could be generated reliably enough that the system becomes usable. Additionally, such a system would be greatly improved, from a usability point of view, if it was

possible for the surface the real flashlight is shone upon, to correspond physically with that depicting the virtual scene it illuminates.

7.3 New Applications

The advantages and opportunities Enlighten presents to the museum and heritage sector are now quite well-understood, as several installations are already in place and being used by the general public on a daily basis. One form of public exhibition that Enlighten has yet to see use in however is the art gallery. A flashlight interface would appear particular suited to use in such environments, as a means to interact with paintings in order to discover more about them. Since paintings are flat, cannot be physically augmented, are sometimes very large and occasionally have audio guide content associated with them (particularly in galleries) their augmentation with flashlights would seem a particularly apt use of the technology.

Enlighten also shows promise for direct use in creative arts, effectively building on the group performances carried out by children who took part in the 'Journey into Space' experience at the Newark Agricultural Show. Continuing the theme of users as content creators (rather than spectators), Dr. Sue Cobb of the University of Nottingham has run a number of informal workshops with music classes at a Nottingham primary school. Here the flashlight system has been used to turn a classroom into an instrument. Using a selection of flashlights, it is possible to play tunes by illuminating different areas. This work represents an interesting variation on the theme of "Music via Motion" proposed by Ng (2002).

The current version of Enlighten may also see use in a number of other, as yet, relatively unexplored areas. The related fields of Special Needs Education, and Rehabilitation are of particular interest. During early commercial development of the system, preliminary interest from SpaceKraft (a company who develop and sell 'sensory rooms' and equipment for use by children with learning difficulties)

in the technology showed potential. Though work reported here has largely concentrated on development of a system for use in museums and heritage, recently Cobb et al (2006, 2007) took Enlighten into the Shepherd School Nottingham. Here, initial trials of the technology were undertaken with a selection of children from different age groups and learning abilities. A system has now been purchased by the school, and special needs teachers are putting it to a variety of uses. Work is underway to develop a commercial product for this sector.

Additionally, Enlighten is likely to find use for rehabilitation, as the 'Motivating Mobility Project' (Rodden 2007), whose goal is to explore the potential application of games and other technology in encouraging stroke suffers to take useful exercise, has also shown interest in using flashlights.

A final area in which Enlighten might find some alternative use is in games or group activities. Through minor alterations to the current system, undergraduate students have already demonstrated a game called 'Ghost Hunter', which sees the player searching for invisible targets using flashlights. Here, audio cues indicate a flashlight's proximity to the wayward target and, additionally, help the player to avoid pitfalls. There is potential that games such as this could be played outdoors at night, using high powered flashlights over larger surfaces, such as the sides of buildings or empty car parks. When utilising these large scale areas, opportunities arise for structured collaboration as a game mechanic. An example of this, in the context of Ghost Hunter, might be to require two players (each with separate flashlights) to track down moving targets simultaneously. Similarly, if multiple cameras were used to cover larger or disjoint areas, a system of logic could be applied which allowed the same flashlight to be recognised in several locations.

7.4 Conclusion

Enlighten has been productively deployed in several museum and heritage attractions that are in daily use, with more installations planned. The technology has been patented in Europe and America (pending) and a successful University spin off company, Visible Interactions Ltd. has been formed to market it. A number of potentially valuable extensions, however, remain to be explored.

7.5 Summary

The focus of this thesis has been on the technical development and deployment of interactive flashlights primarily for individual use in the museums and heritage sector. Some technical extensions to the techniques currently employed are possible and, with further work, the method might be extended to different physical situations for example, the use of flashlights to illuminate virtual environments. Additionally, the interface is also showing promise for use in a number of other applications outside of museum and heritage sectors, the most notable of these being in special needs and education. The method has been patented and a spin off company has been formed to market it.

Appendix A

The Enlighten User Interface

Conceptually, Enlighten is very different from the environment which it was built on (CamCap) however utilisation of the underlying technology remains the same. As noted in Chapter 4, the requirements of a deployable system primarily revolve around the need for full (hands off the interface) automation, customisation to the specific application and an easy to use, logical and intuitive interface. Additionally, a greater level of customisation is required than what is possible with CamCap (see Appendix B) together with the ability to open or limit this in accordance with customer needs or purchase privileges.

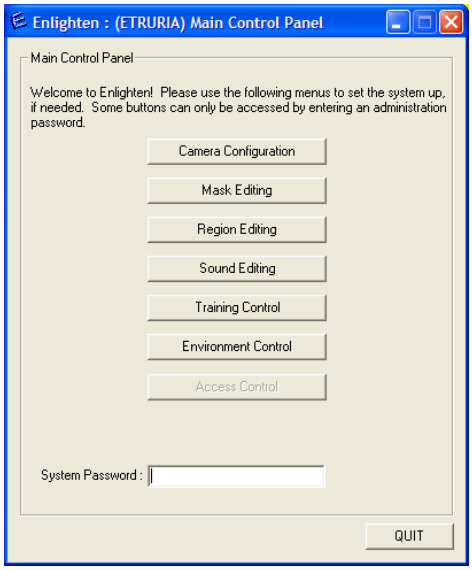


Fig A.1 - Enlighten Main Interface

The program is therefore organised as a wizard style application with stages of configuration presented in order (see figure A.1), each having multiple pages to allow separate settings for the individual cameras used (up to a maximum of ten). In Enlighten, the "stages" of configuration consist of the following as listed below, the latter being a security feature providing the means to allow or restrict access to system settings once an initial configuration has been established.

Camera Configuration

Allowing the selection and naming of cameras together with configuration of properties such as frame rate, exposure etc (Fig. A2i)

Mask Editing

An art-package paintbrush style interface allowing the masking of areas from a camera's field of view. For example, a corridor beside a planar wall (Fig. A2ii)

Region Editing

Box-drag style interface for defining regions (targets) in a camera's field of view for Enlighten to associate content with (Fig. A2iii)

Sound Editing

Allowing association and previewing of up-to ten sounds per target (Fig. A2iv) (Note: At Magna a customised version triggered midi signals instead)

Training Control

An interface governing control of flashlight training in each camera's field of view allowing recognition to trigger different content (Fig. A2v)

Environment Control

Optional settings and parameters allowing fine tuning of the system (Fig. A2vi) (see Chapter 5 for details)

Access Control

Password protection to allow access to all or some of the above stages of configuration by end users

While some of these settings existed in previously discussed custom versions of CamCap (for example definition of active regions and sound allocation in the cave system described in Chapter 2), in Enlighten, the interface was significantly improved to not only allow for configuration on a keyboard-less touch screen system (effectively eliminating the ability to "left click") but also ease of target resizing and repositioning for example without need for redefinition. Aside from benefits to the usability of the system, such enhancements, when combined with the noted option to disable access to certain stages of configuration, meant that it was possible to deploy a system that allowed alterations such as repositioning or

resizing of audio targets (for example to account for damaged exhibits or camera shift) without permitting the creation of actual new content. Together with allowing Enlighten to be licensed at varying levels of customisation, such restrictions can also be used in order to avoid any potential damage to a configured system by inexperienced users.

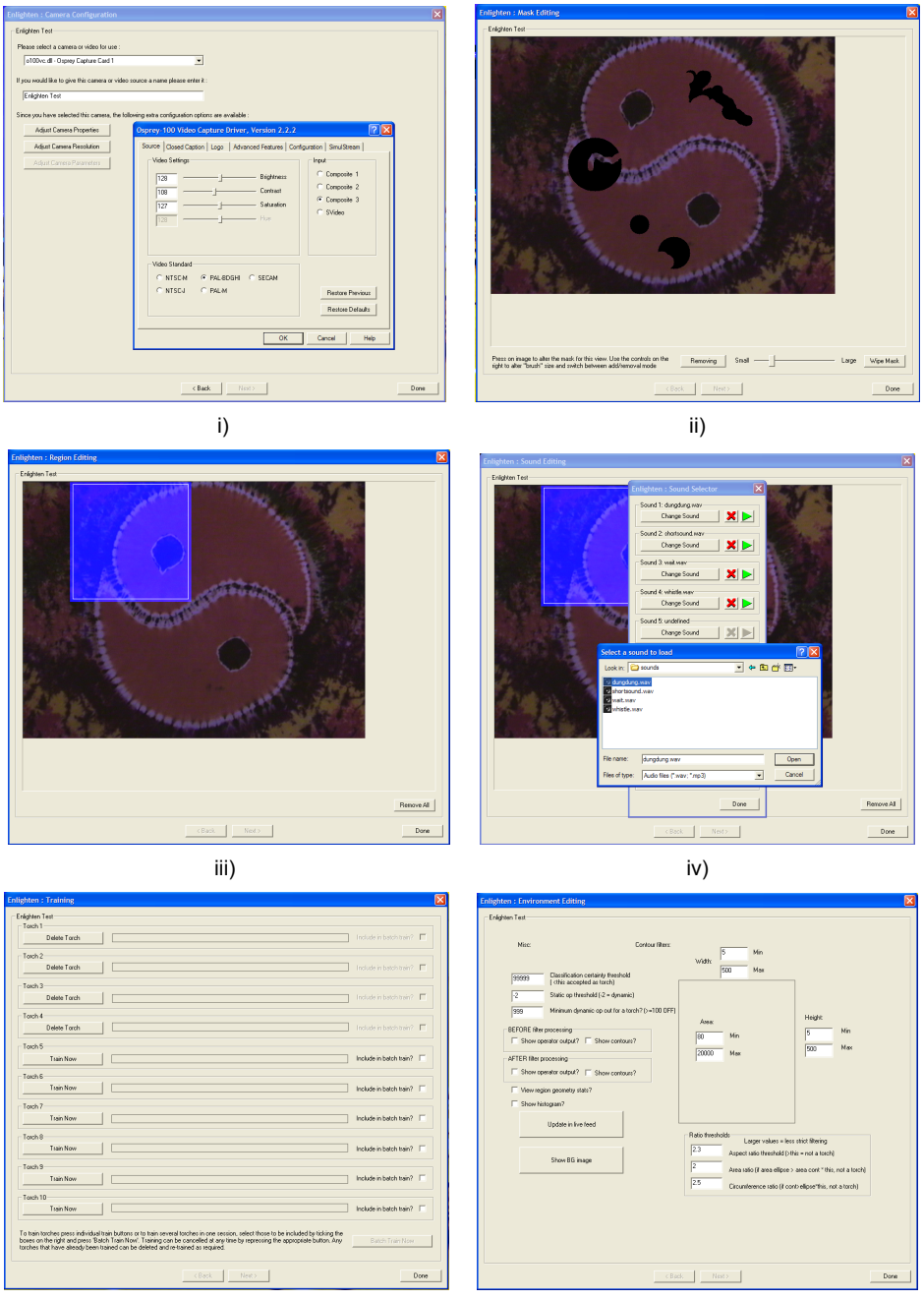


Fig A.2 – Enlighten Configuration Screens: i) Camera Configuration ii) Mask Editing iii) Region Editing iv) Sound Editing v) Training Control vi) Environmental Control

Appendix B

Underlying Technology: CamCap

The Enlighten software is based on another title, written during an early stage of software development: CamCap. CamCap is a computer vision programming and prototyping environment available from the author's website (NottsVision). It was constructed to provide easy access to, process and work with the raw data within any video stream (regardless of compression format), single image, or CCD camera output. Video capture can be achieved via a standard capture card or by interfacing directly to commercial webcams or firewire cameras.

Although other environments, that provide similar functionality exist (e.g. MatLab, Wit, Khoros) CamCap is particularly useful when converting prototype software into commercial applications that are compatible with a client's hardware (as was the case with Enlighten).

CamCap's strengths lie within the fact that it provides, ready to use, much of the base programming typically required for a video processing application. The software makes extensive use of the Windows and DirectX API's, Microsoft Foundation Classes and Component Object Model (MFC and COM) and the intel open-source computer vision library (OpenCV). The program is designed and assembled in a manner that enables the user to have little or no understanding of such technologies, yet still be able to benefit from their use. Additionally, by utilising the OpenCV image format CamCap not only provides the basics but also support for a wide range of classic and current image processing methods and techniques (intel OpenCV SDK). These are all optimized for use with a variety of different intel manufactured chipsets from Pentium II and up.

CamCap has been extensively used by undergraduate and postgraduate students as a springboard for work on image/video processing based projects without having

to undergo a steep learning curve. It has also been utilized as a teaching resource for vision based modules the world over.

In its distributed format, CamCap consists of a basic dialog based application, with two sets of mirrored controls allowing the loading, transport and display of two video streams/camera captures or loaded/captured images. For video files, functionality is provided for play, stop, forward, rewind, frame advance/reverse and speed control together with region selection and grabbing of either modified or unmodified frames from the stream. When cameras are used as the video source, the program exposes customised graphical interfaces, allowing alteration of properties such as frame rate, resolution and shutter speed, dependant the on functionality exposed by their drivers. With minimal MFC knowledge, further customisation of the application is easily achieved by adding processing routines to the examples provided in drop down menus on either side of the program (Fig. B.1). Video processing, at its most basic level, can be achieved by the manual capture of frames, from the video stream, and executing the code associated with each of these tailor made functions.

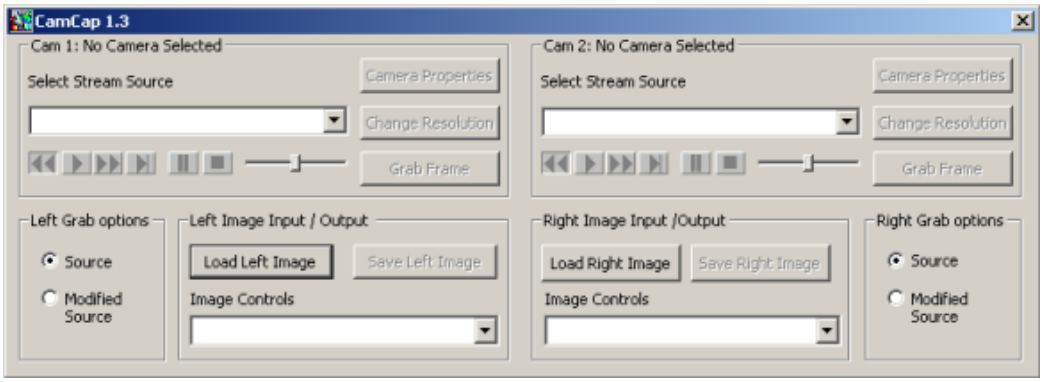


Fig B.1 – CamCap Control Interface

Aside from the "frame-capture and process" programming model described above, video processing is encouraged "in situ" for efficiency and, to this end, customisable functions are provided for each of the two video streams available. These are called on a frame by frame basis exposing data from the stream in the OpenCV image format, so as to allow for its easy manipulation using the range of functions provided by this library.

To achieve this, CamCap utilizes the Direct X graph model whereby a conceptual flow of data is constructed using filters to represent each stage in the process of rendering a video stream. Typically, such a stream can be built automatically, starting from file or camera source, through to the insertion of appropriate decompression filters, colour space converters and, finally, an appropriate rendering technique (Fig. B.2). Under usual circumstances, the only way of modifying data within the stream is the programming of a custom made filter and its manual insertion into the graph. CamCap however utilizes what is, in effect, a proxy filter inserted just before the rendering stage. This enables call outs to the aforementioned external functions which contain the frame data and this can be processed either in place or in a user defined external class. Use of such a model, which is of comparable computational efficiency to writing custom filters, is the key to the rapid prototyping and flexibility provided by the platform. In addition to this, the pre-written support for dual playback of video streams allows for concurrent access of data from either stream simultaneously, thus providing the necessary framework required for stereo processing, should such functionality be desirable.

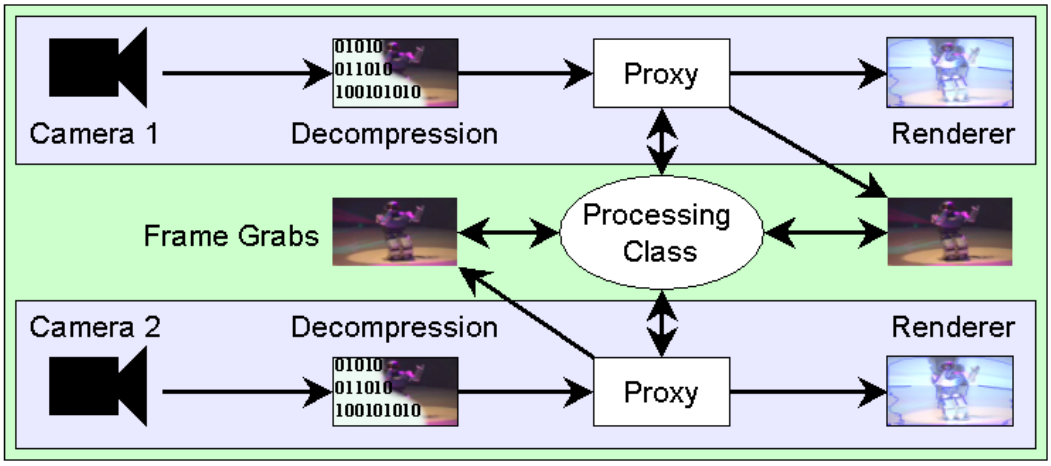


Fig B.2 – Data Stream Rendering and Access in CamCap

In summary, the key features of CamCap are:

- Supports all video capture devices from frame grabbers to web cams whose drivers support either DirectShow or the older Video for Windows (VFW) format

- Handles most popular video compression formats (mpeg, indeo, divx, xvid, asf etc) for processing of video files.

- Load, display and save image files from/in a variety of formats.

- Full transport control of video files: play, stop, end, pause and frame advance/backup.

- Variable playback speed control for video files from $1/10^{th}$ speed to double.

- Frame rate control for capturing from devices.

- Capture, process[1] and save both source and modified frames from video streams.

- Built in static image processing functions including region selection and image marking and a variety of other common methods.

- Allows easy addition of custom processing functions accessible via drop down menus.

- Promotes fast development of Computer Vision algorithms using video processing without knowledge of Microsoft MFC[2], Direct X or COM programming.

- Simultaneous processing of two video streams (devices or files).

- Provides simple implementation of stereo algorithms.

- Modular design for easy extension and modification.

- Uses Microsoft Visual Studio Development Environment.

- Full support for the intel Open Source Computer Vision libraries' image format.

[1] In addition to video processing using prewritten image operators
[2] Only required if additional interface controls are required

Appendix C

Summary of Collaborations

In this thesis, many of the described installations or experiments represented significantly large projects and hence were worked on by several individuals. In the StoryTent, Dr. Borianna Koleva carried out the 3D programming side of the installation whereas Dr. Holger Schnadelbach was responsible for the tent construction and RFID tag reader code; the original flashlight interface was programmed by Jonathan Green. The 3D design and network programming used in the StoryTent was adapted by Dr. Mike Fraser to create the Sandpit which, together with the Caves installation, was set up, run and analysed with additional input from Jonathan Green, Dr. Sahar Bayomi, Dr. Andrew French, Prof. Steve Benford and Dr. Tony Pridmore. The flashlight interface deployed in the caves was adapted by Dr. Ahmed Ghali (from that used in the StoryTent) and a new version was produced for the MRL and Newark show installations by Jonathan Green. The 'Journey into Space' drama experience, used during the Newark show, was designed and run by Rachel Fenely with assistance from Jonathan Green, Dr Tony Pridmore and Stuart Reeves

Enlighten was written by Jonathan Green and initially used at Etruria, Intech, and MAGNA then later, by Dr. Sue Cobb at Shepherd School. The Etruria Installation was carried out by Jonathan Green and Dr. Andrew French with assistance from Dr. Tony Pridmore, Dr. Tony Glover and Dr. Sue Cobb. Both Intech and MAGNA installations were completed by their own staff in collaboration with Dr. Tony Glover, Prof. Steve Benford, Dr. Tony Pridmore and Jonathan Green. Additionally, the Intech and MAGNA versions of Enlighten were customised by Jonathan Green and Dr. Tony Glover who also redesigned the system's graphical user interface for use in MAGNA's planned 'second phase' installation.

Early experiments on recovering rotational orientation from flashlight projections resulted from a collaboration between Jonathan Green and Dr. Steve Mills.

References

Aoki P., & Woodruff A., Improving Electronic Guidebook Interfaces Using a Task-Oriented Design Approach. *DIS 00, 319-315, Brooklyn, New York*

Aleksander I. and Burnett P., Thinking Machines. *p126-127, 1987*

Amtel Corp – High Dynamic Range Cameras:
http://www.atmel.com/dyn/resources/press/akyla.html

Bailey T., Jain A. A note on distance-weighted k-nearest neighbor rules. *IEEE Trans. Systems, Man, Cybernetics, Vol. 8, pp. 311-313, 1978*

Barsky S.and Petrou M., The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE PAMI, 25(10):12391252, 2003*

Bayon V., Wilson J.R., Stanton D., Boltman A., Mixed Reality Storytelling Environments. *Virtual Reality, Volume 7, 2003*

Blake A., Boundary Conditions for Lightness Computation in Mondrain World. *Computer Vision, Graphics and Image Processing 314-327, 1985*

Bobick A., Intille S., Davis J. et al, The KidsRoom:A perceptually based interactive and immersive storyenvironment. *PRESENCE: Teleoperators and Virtual Environments 8, 4 (Aug. 1999), 367–391*

Bradski G. and J. Davis, Motion Segmentation and Pose Recognition with Motion History Gradients. *Machine Vision and Applications, Vol. 13, No. 3, pp. 174-184, 2002*

Brelstaff G. and Blake A., Detecting Specular Reflections Using Lambertian Constrains. *2nd International Conference on Computer Vision, pp 297-302, 1988*

Cobb S., Mallet T., Pridmore T. and Benford S., Interactive Flashlights in Special Needs Education. *ArtAbilitation, Esbjerg, Denmark, 2006*

Cobb S., Mallet T., Pridmore T. and Benford S., Interactive Flashlights in Special Needs Education. *Digital Creativity, 18(2), 69-78, 2007*

Colombo C., Del Bimbo A. and Valli A., Visual capture and understanding of hand pointing actions in a 3-D environment. *IEEE Transactions on Systems, Man, and Cybernetics, vol. 33 (4), pp. 677-686, 2003*

Comaniciu D., Ramesh V., and Meer P., Kernel-based object tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence, 25(5), p. 564–577, 2003*

Cootes T., Taylor C., Cooper D., Graham J., Active Shape Models – Their Training and Application. *Computer Vision and Understanding, Volume 61, p38-59, 1995*

Craven M., Taylor I., Drozd A., Purbrick J., Greenhalgh C., Benford S., Fraser M., Bowers J., Jää-Aro K., Lintermann B., Hoch M., Exploiting interactivity, influence, space and time to explore non-linear drama in virtual worlds. *CHI 2001 Proc pp 30-37, ACM Press*

Davis J. & Chen X., LumiPoint: multi-user location-based interaction on large tiled displays. *Displays, Vol. 23-5, Elsevier Science 2002*

Davis J., and Vaks S., A perceptual user interface for recognizing head gesture acknowledgements *Proceedings of the 2001 ACM Workshop on Perceptive User Interfaces*, *Orlando, 2001*

Denman H., Rea N., Kokaram A., Content-based analysis for video from snooker broadcasts *Computer Vision and Image Understanding. Volume 92, Issues 2-3, 2003*

Debevec P., Malik J., Recovering High Dynamic Range Radiance Maps from Photographs. S*IGGRAPH, Annual Conference Series, ACM, 369–378, 1997*

Dudani, S.A. The distance-weighted k-nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern., SMC-6:325–327, 1976*

Dunn M. E., Joseph S. H., Processing Poor Quality Line Drawings by Local Estimation of Noise. *Proc. 4th International Conference on Pattern Recognition, p153-162, 1988*

El Kaliouby R. and Robinson P., Real-time head gesture recognition in affective interfaces. *Proceedings of the 2003 IFIP TC13 International Conference on Human-Computer Interaction*, pp. *950-953, 2003*

Elrod S., Bruce R., Gold R., Goldberg D., Halasz F., Janssen W., Lee D., McCall K., Pedersen E., Pier K., Tang J., and Welch B., LiveBoard: A large interactive display supporting group meetings, presentations and remote collaboration. *CHI'92 Proceedings, pp. 599-607, 1992*

Fix E., Hodges J., Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties. *California University Berkeley, 1951*

Forbus K., Light Source Effects. *A.I. Memo 422, A.I. Lab, M.I.T, 1977*

Freeman W. T, Adnerson D., Beardsley P., Dodge C., Kagel H., Kyuma K., Miyake Y., Roth M., Tanaka K., Weissman J. C. and Yerazunis W., Computer Vision for Interactive Computer Graphics. *IEEE Computer Graphics and Applications, May-June, pp. 42-53, 1998*

Freeman W. T., Beardsley P. A., Kahe H., Kyuma K. and Weissman C. D., *SIGGRAPH Computer Graphics Magazine, TR-99-36, November 1999*

Freeman W. T, Tanaka K., Ohta J. and Kyuma K., Computer vision for computer games. *Mitsubishi Electric Advanced Technology R&D Center 8-1-1, Tsukaguchi-Honmachi Amagasaki City, Hyogo 661, Japan*

Freeman W. T. and Weissman C. D., Television control by hand gestures. *IEEE Intl. WkShp on Automatic Face and Gesture Recognition, Zurich, 1995*

Gao D., Zhou J., Xin L., SVM-based detection of moving vehicles for automatic trafficmonitoring. *Intelligent Transportation Systems, 2001*

Gorodnichy D.O., Malik S., Roth G., Nouse: Use your Nose as a Mouse - a New Technology for Hands-free Games and Interfaces. *15th Int.l Conf. on Vision Interfaces Calgary, Canada, May 2002*

Green J., Pridmore P., Benford S., Ghali A., Location and Recognition of Flashlight Projections for Visual Interfaces. *Proc., 17th International Conference on Pattern Recognition (ICPR'04) Volume 4, 2004*

Green J., Schnädelbach H., Koleva B., Benford S., Pridmore T., Medina K., Harris E., Smith H. Camping in the digital wilderness: tents and flashlights as interfaces to virtual worlds. *Chi 2002 Ext. Abs. pp 780-781, 2002*

Greenhalgh, C. & Benford, S. MASSIVE: a Distributed Virtual Reality System Incorporating Spatial Trading Proc. *IEEE 15th International Conference on Distributed Computing Systems (DCS'95), Vancouver, Canada, May 30 - June 2, 1995*

Greenhalgh C., Izadi S., Rodden T., Benford S.. The EQUIP Platform: Bringing together Physical and Virtual Worlds. *Technical Report, University of Nottingham, 2001*

Ghali G., Bayomi S., Green J, Pridmore T., and Benford S., Vision-mediated Interaction with the Nottingham Caves. *Proceedings SPIE Conf. on Computational Imaging, 2003*.

Ghali A., Benford S., Bayoumi S., Green J., Pridmore T., Visually-tracked Flashlights as Interaction Devices, *Interact 2003*

Glasby E. An analysis of histogram-based thresholding algorithms. *Graphical Models and Image Processing: 55; 6. 1993*

Grimson W.E.L., Stauffer C., Romano R. & Lee L. Using adaptive tracking to classify and monitor activities in a site. *Proc. IEEE Computer Vision and Pattern Recognition* 22-31*, 1998*

Heap A.J., Real-time hand tracking and gesture recognition using smart snakes. *Proc. Interface to Real and Virtual Worlds, 1995*.

Hongshen M. and Paradiso J., The FindIT Flashlight: Responsive Tagging Based on Optically Triggered Microprocessor. *Proc. Ubicomp, 2002, Gothenberg, Setpember, 2002*

Horn, B.K.P., Determining Lightness from an Image. *Computer Graphics and Image Processing, Vol. 3, No. 1, December 1974*

Intel Open Source Computer Vision SDK: http://www.intel.com/research/mrl/research/opencv/

Isard, M. and Blake, A., Condensation - conditional density propagation for visual tracking, *International Journal of Computer Vision, vol. 29, no. 1, pp. 5-28, 1998*

Joseph S.H., Processing of line drawings for automatic input to CAD. *Pattern Recognition 22: 1-11, 1989*

Kalman R.E., A New Approach to Linear Filtering and Prediction Problems Trans. *ASME - Journal of Basic Engineering 82 (Series D), 35-45, 1960*

Kanbara M. and Yokoya N., Real-time Estimation of Light Source Environment for photorealistic Augmented Reality. *17th International conference on Pattern Recognition, ICPR ' 04, 2004*

Kang H., Lee C.W. and Jung K., Recognition-based gesture spotting in video games. *Pattern Recognition Letters, vol. 25 (15), pp. 1701-1714, 2004*

Kapoor A. and Picard R. A real-time head nod and shake detector. *Proceedings of the 2001 ACM Workshop on Perceptive User Interfaces, 2001*

Kato H., M. Billinghurst, I. Poupyrev K. Imamoto K. Tachibana, Virtual Object Manipulation on a Table-Top AR Environment. *Proceedings of ISAR 2000, Oct 5th-6th, 2000*

Kawato S. and Ohya J. "Real-time detection of nodding and head-shaking by directly detecting and tracking the 'between-eyes'", *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 40-45, *2000*

Khan S.and Shah M., A Multiview Approach to Tracking People in Crowded Scenes Using a Planar Homography Constraint. *ECCV, 2006*

Khoros: http://www.khoral.com/

King H., An Investigation into Dynamically Lighting a Virtual World with and Real Light Source. *Unpublished Undergraduate Dissertation, University of Nottingham, 2006*

Kittler J. and J. Illingworth, Minimum Error Thresholding. *Pattern Recognition 19: 41-4, 1986*

Koleva B., Schnädelbach H., Benford S. & Greenhalgh C., Traversable Interfaces Between Real and Virtual Worlds. *CHI'2000, 2000*

Lambert J. H., Photometria sive de mensure de gratibus luminis, colorum umbrae. *Eberhard Klett, 1760*

Land E. H., McCann J. J, Lightness and Retinex Theory. J*ournal of the Optical Society of America, 61, 1-11, 1971*

Marks R., EyeToy: A New Interface for Interactive Entertainment. *Computer Systems Laboratory Colloquium, 2004*

MATLAB: http://www.mathworks.com/products/matlab/

Morimoto C., Koons D., Amir A. and Flickner M. Real-time detection of eyes and faces. *Proc. Workshop on Perceptual User Interfaces pp. 117-120, 1998*

Myers B., Bhatnager R., Nichols J., Peck C., King D., Miller R. & Long C., Interacting at a distance: measuring the performance of laser pointers and other devices. *CHI 2002, 33-40, Minneapolis, USA, April 2002, ACM*

National Trust (the): http://www.nationaltrust.org.uk/main/w-trust/w-thecharity.htm

Nesi P. and Del Bimbo A., A vision-based 3-D mouse. *International Journal of Human-Computer Studies*, *vol. 44 (1), pp. 73-91, 1996*

Newark and Nottinghamshire County Show: http://www.newarkshowground.com/cs_home.htm

Ng K., Sensing and mapping for interactive performance. *Organised Sound 7, 2, pp 191-200, 2000*

Ng. Kher Hui. Mixed Reality Interaction for Group Experience, *2002*

Nillius P. and Eklundh J., Automatic Estimation of the Projected Light Source Direction, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPA ' 01), 2001*

NottsVision: http://www.cs.nott.ac.uk/~jzg/nottsvision/

Okatani T. and Deguchi K., Estimation of Illumination Distribution Using a Specular Sphere, *15th International Conference on Pattern Recognition (ICPR ' 00), 2000*

Olsen D., Nielsen T., Laser Pointer Interaction, *CHI 2001Proc. 17-22, 2001*

Opera Software – Mouse Gestures Documentation:
http://www.opera.com/products/desktop/mouse/

Pirhonen A., Brewster S.and Holguin C., Gestural and Audio Metaphors as a Means of Control for Mobile Devices. *CHI 2002 Proc pp 291-198. ACM Press, 2002*

Proesmans M., Van Gool L., Pauwels E. and Oosterlinck A., Determination of optical flow and its Discontinuities using non-linear diffusion. *3rd Eurpoean Conference on Computer Vision, ECCV'94, Volume 2, pages 295-304, 1994*

Ramey-Gassert L., Walberg III H.J., Walberg H.J. Re-examining connections: Museums as science learning environments. *Science Education, 78(4), 345-363, 1994*

Regenbrecht H., Wagner M., Interaction in a Collaborative Augmented Reality Environment. *CHI 2002 Extended Abstracts, 504-505, 2002*

Rehg J. and Kanade T., DigitEyes: Vision-Based Human Hand Tracking. *Proc. European Conference on Computer Vision, 1994*

Rekimoto J., Matushita N., Perceptual Surfaces: Towards a Human and Object Sensitive Interactive Display. *Workshop on Perceptual User Interfaces (PUI'97), 1997*

Rodden T., Mawson S., Probert Smith P., Ricketts I., Burridge J., Fitzpatrick G., Motivating Mobility: Interactive Systems to promote Physical Activity and Leisure for people with limited mobility. *EPSRC Project (Reference EP/F00382X/1), 2007*

Rosin P.L., Unimodal Thresholding. *Proc. Scandanavian Conference on Image Analysis, pp. 585-592, Kangerlusuuaq, Greenland, 1999*

Rosin, P.L., Unimodal Thresholding. *Pattern Recognition, vol. 34, no. 11, pp. 2083-2096, 2001*

Sahoo P.K. et al. A survey of thresholding techniques. *Computer Vision, Graphics and Image Processing: 41; 233-260. 1988*

Schnädelbach H., Koleva B., Flintham M., Fraser M., Izadi S., Chandler P., Foster M., Benford S., Greenhalgh C., Rodden T.. The Augurscope: A Mixed Reality Interface for Outdoors. *Proc. CHI 2002 pp 9-16. ACM Press, 2002*

Schnädelbach H, Penn, A., Benford S., Steadman P., Koleva, B., Moving Office: Inhabiting a Dynamic Building. *CSCW conference, Banff, Canada, pp.313-322, 2006*

Sloan Digital Sky Survey: http://www.sdss.org

Sonka, Hlavac and Boyle, Image Processing, Analysis and Machine Vision. *p80-81, 1994*

SpaceKraft – Interactive Products and Multisensory Environments for Special Needs: http://www.spacekraft.co.uk

Stanton D., Bayon V., Neale H., Ghali A., Benford S., Cobb S., Ingram R., O'Malley C., Wilson J., Pridmore T., Classroom Collaboration in the Design of Tangible Interfaces for Storytelling. *CHI 2001 Proc pp 482-489, ACM Press, 2001*

Stauder J., Mech R., Ostermann J., Detection of Moving Cast Shadows for Object Segmentation. *IEEE Transactions on Multimedia. 1999*

Stewart J., Bederson B., & Druin A., Single display groupware: A model for co-present collaboration. *CHI99, pp. 287-288, 1999*

Tang W. and Rong G. A real-time head nod and shake detector using HMMs. *Expert Systems with Applications*, vol. 25, pp. 461-466, 2003

Trucco E., Verri. A., Introductory Techniques for 3-D Computer Vision. *P56-60, 1998*

Tsukiyama T., Inferring the 3D shape formed by plane surfaces from isoluminance curves. *Image and Vision Computing Volume 13 Number 9, 1995*

UK Patent Publication Number GB2411957 – Optically triggered interactive apparatus and method of triggering said apparatus: http://www.wikipatents.com/gb/2411957.html

Ullman S., On Visual Detection of Light Sources. *Biology and Cybernetics 21, pp 205-212, 1976*

United Kingdom Infrared Deep Sky Survey: http://www.ukidss.org

US Patent Application Number 10/540,498 – Optically triggered interactive apparatus and method of triggering said apparatus: http://www.freepatentsonline.com/y2007/0008279.html

Viola P., Jones M., Robust Real Time Object Detection. *2nd Interenational Workshop on Statistical Learning and Computational Theories of Vision – Modeling, Learning, Computing and Sampling, 2001*

Waterworth J., and Waterworth E., In Tent, In Touch: beings in seclusion and transit. *CHI 2001 Ext. Abs, pp 303-304, 2001. ACM Press, 2001*

WiT: http://www.logicalvision.com/

Woodham R. J., Photometric method for determining surface orientation from multiple images. *Optical Engineering, 19(1):139144, 1980*

vom Lehn D., Heath C. & Hindmarsh J., Exhibiting Interaction: Conduct and Collaboration in Museums and Galleries. *Symbolic Interactio*n, 24 (2), *University of California Press, 2002*

Zhou W. and Kambhamettu C., Estimation of the Size and Location of Multiple Area Light Sources. *17th International Conference on Pattern Recognition (ICPR ' 04),* 2004

Zhou W. and Kambhamettu C. Estimation of Illuminant Direction and Intensity of Multiple Light Sources., *ECCV 2002*, *2002*