# Depth Measurement in Integral Images

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ChunHong WU

3D Imaging Group
Faculty of Computer Sciences and Engineering
DE MONTFORT UNIVERSITY, UK

February 28, 2003

# Abstract

The development of a satisfactory the three-dimensional image system is a constant pursuit of the scientific community and entertainment industry. Among the many different methods of producing three-dimensional images, integral imaging is a technique that is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution by using unique optical components. The generation of depth maps from three-dimensional integral images is of major importance for modern electronic display systems to enable content-based interactive manipulation and content-based image coding. The aim of this work is to address the particular issue of analyzing integral images in order to extract depth information from the planar recorded integral image.

To develop a way of extracting depth information from the integral image, the unique characteristics of the three-dimensional integral image data have been analyzed and the high correlation existing between the pixels at one microlens pitch distance interval has been discovered. A new method of extracting depth information from viewpoint image extraction is developed. The viewpoint image is formed by sampling pixels at the same local position under different micro-lenses. Each viewpoint image is a two-dimensional parallel projection of the three-dimensional scene. Through geometrically analyzing the integral recording process, a depth equation is derived which describes the mathematic relationship between object depth and the corresponding viewpoint images displacement. With the depth equation, depth estimation is then converted to the task of disparity analysis. A correlation-based block matching approach is chosen to find the disparity among viewpoint images.

To improve the performance of the depth estimation from the extracted viewpoint images, a modified multi-baseline algorithm is developed, followed by a neighborhood constraint and relaxation technique to improve the disparity analysis. To deal with the homogenous region and object border where the correct depth estimation is almost impossible from disparity analysis, two techniques, viz. Feature Block Pre-selection and “Consistency Post-screening, are further used. The final depth maps generated from the available integral image data have achieved very good visual effects.

# Acknowledgements

My PhD study has been a pleasant experience and this thesis brings that journey to an end. I am deeply indebted to Prof. Malcolm McCormick and Dr. Amar Aggoun, my supervisors at De Montfort University, who have been the most important figures in making my three years here so rewarding. I am very grateful to Prof. SunYuan Kung, my advisor at Princeton University in the USA, for his advice and guidance. I hope that this thesis can partly repay them for their persistence and dedication to this cause.

I would like to extend my thanks to the members of the 3D Imaging Group, the support staff at De Montfort University for their assistance and co-operation throughout the project. I would also like to acknowledge the support and help provided by the Department of Electronic Engineering at Princeton University during my stay in the Department.

De Montfort University is gratefully thanked for providing my full PhD Scholarship without which this work would not have been possible.

My life in the UK was enriched by many new friends, and I would like to thank them without going into a long list of individual names.

I am greatly indebted to my parents and sisters, in particular my eldest sister and brother-in-law, for their love, encouragement and unfailing support. My special thanks must go to my husband, without his understanding and supporting, my PhD course would not have been possible. It is my intention to dedicate this thesis to them to express my deep indebtedness and sincere gratitude.

ChunHong Wu,

Leicester, UK, January 2003

# Statement of Originality

The work contained within this thesis is purely that of the author unless otherwise stated.

ChunHong Wu

# Acronyms

| | |
|---|---|
| 3D | Three-dimensional |
| 3DTV | Three-dimensional Television |
| CCD | Charge Couple Device |
| DMU | De Montfort University |
| II | Integral Image |
| VI | Viewpoint Image |
| VIP | Viewpoint Image Pair |
| UII | Unidirectional Integral Image |
| OII | Omnidirectional Integral Image |
| LCD | Liquid Crystal Display |
| psf | Point Spread Function |
| PSF | the matrix associated with Point Spread Function |
| SFM | Structure from Motion |
| ATS | Auto-collimating Transmission Screen |
| DVD | Digital Video Disk |
| MPSF | Modified Point Spread Function Matrix |

# Contents

# List of Figures

# List of Tables:

# List of Integral Images

# Chapter 1

# Introduction

## 1.1 The research area

The development of three dimensional (3D) imaging systems is a constant pursuit of the scientific community and entertainment industry (Motoki1995, Taub2002, 3dcgi). There is growing evidence that 3D imaging techniques will have the potential to establish a future mass-market in the fields of entertainment and communications. One of the much discussed applications that exist for 3D display and video communication systems is 3D television.

Many different approaches have been adopted in attempts to realize free viewing 3D TV systems (Okoshi1976, McAllister 1993). Holographic systems can produce still images with high quality and give an extremely realistic reproduction of the spatial images but have difficulty in producing and displaying moving spatial images due to the requirements for the coherent light sources (Gabor1948, Leith1963, Outwater1999). Recently, several groups have demonstrated autostereoscopic Television systems (Actualdepth, DTI, DDD, 4D-Vision, Philips, Genex, Stereographics). Most of them work on the principle of presenting multiple images to the viewer by use of temporal or spatial multiplexing of several discrete viewpoints to the eyes. The viewing effect depends on the viewpoint number, a higher quality experience being enjoyed when the viewpoint number is high. This creates the problem of simultaneously generating or capturing enough views in real time at an affordable cost which cause major difficulty in the displaying system. To date, integral imaging is a technique that is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution by using unique optical components (McCormick 1995, Okano 1998, Min 2001, Naeumra 2001, Yano 2002).

An innovative and unique optical system for transferring full parallax three dimensional images has been developed by the 3D Imaging Group at De Montfort University (DMU) (Davies 1994, McCormick1995). The two-tier optical camera arrangement overcomes the image degradation caused by the two-stage recording process and allows direct spatially correct 3D image capture for orthoscopic display. Subsequently, Okano et al reported an integral GRINROD optical transmission array that produced 3D image data suitable for imaging by 4 multiplexed CCD elements, an arrangement previously proposed to achieve the required resolution by McCormick et al (Arai 1998, McCormick 1992). To date most researchers have concentrated on establishing appropriate viewing parameter characterization and improved image generation. In respect of stereoscopic systems a number of groups have tackled data compression and computer graphical generation of image views (Min 2001, Naeumra2001). However there are many data processing issues that require specialist solutions unique to integral image (II). One of these issues, the knowledge of spatial position is particularly useful to enable content-based interactive manipulation and content-based image coding. The work in this thesis addresses the particular issue of analyzing integral images in order to extract spatial position information from the planar recorded II.

## 1.2 Scope of the thesis and original contribution

An integral image has a unique image format. In the simplest form, the image is captured by a special camera having a photographic emulsion placed behind a micro-lens array whose back surface is coincident with lens focal length. The 3D information from the object space is embedded in two-dimensional (2D) recording data. Although the data exists in 2D format, it contains all the 3D spatial information that can be replayed to reconstruct a true 3D optical model.

To find out the way of extracting depth information from the integral images, two aspects are investigated: the general depth measurement methods used in the computer vision area and the mechanism of spatial information encoding within integral images. Previously, a mathematical model of the integral imaging system has been analyzed and described and an approach to the extraction of depth information based on the image inverse theory has been investigated (Manolache 2001, 2002). The approach uses the point-spread function of the optical recording to describe the associated integral imaging system and tackles the 3D spatial reconstruction task as an inverse problem. However, the image inverse problem proves to be ill posed and the discrete correspondents are ill conditioned. The approach can only work on the simulation due to the inherent loss of information associated with the model in the recording process. Analysing the mechanisms existing in the spatial encoding of the II data, current work leads to a new approach for obtaining depth information through the formation of viewpoint images.

Each viewpoint image is extracted utilising the unique optical properties associated with the recording of the integral image. The viewpoint image is a 2D parallel recording of the 3D space and is significantly different from the traditional 2D image taken with an ordinary camera. How the viewpoint image is formed is explained in the thesis. A number of possible applications of viewpoint image extraction are given.

Following viewpoint image extraction, a mathematical relationship, which describes the depth of an object point and the corresponding recording position in two viewpoint images, is derived through geometrical analysis of the II recording process. A depth equation is then formed to give the relationship between the depth and the corresponding displacement between two viewpoint images. Therefore, the task of depth estimation is converted to the task of disparity analysis among viewpoint images. The disparity analysis techniques developed by other researchers are adapted according to the unique format of the Integral image to improve the performance of depth estimation. The hybrid algorithm constructed consists of a modified multi-baseline algorithm, neighbourhood constraint and relaxation technique, multi-candidates pre-screening technique, feature block pre-selection technique and consistency post-screening technique.

The thesis contains the following original contributions:

1. A new way of analyzing 3D integral imaging system through viewpoint image extraction.
2. A new depth equation which gives the mathematical relationship of the object depth and the displacement of the  corresponding viewpoint images.
3. A novel way of obtaining depth information from the II through viewpoint image extraction and disparity analysis.
4. A modified multi-baseline algorithm used to improve the depth estimation performance from the extracted viewpoint images.
5. Application of the neighbourhood constraint and relaxation techniques to improve the depth estimation performance from the extracted viewpoint images.
6. Application of the feature block pre-selection and consistency post-screen technique to improve the depth estimation performance from the extracted viewpoint images.
7. Achievement of accurate depth measurement result and acceptable depth maps from realistic 3D II.

## 1.3 Outline of the thesis

Chapter1 introduces the subject of the research work, providing an overview of 3D integral image, the need for depth information in 3D integral image processing and the scope and original contribution of this work.

Chapter2 contains a brief historical overview of two 3D imaging techniques: stereoscopic and autostereoscopic display.  It concentrates on the free view (autostereoscopic) system and provides a detailed description of the 3D integral imaging system developed by the 3D Imaging Group at De Montfort University.

Chapter3 begins to address the task of depth extraction from 3D integral images. Common depth extraction methods in the computer vision area and previous work on depth extraction from integral image (II) are investigated.

Chapter4 presents a new approach to the extraction of the depth information through viewpoint image extraction. The unique characteristic of 3D integral image data is analysed, and the direction selectivity existing in integral recording is observed. A depth equation, which gives the object depth and corresponding displacement between viewpoint images, is then derived through geometric analysis of the integral recording process. After viewpoint image extraction, a correlation-based block matching method is used to obtain the disparity between viewpoint images. To measure the accuracy of the depth estimation, a simple object scene which only contains two depth levels is specially designed. The depth measurement result is given in this chapter.

In Chapter5, a modified multi-baseline algorithm with neighbourhood constraint and relaxation technique is adapted to improve the depth measurement performance on the extracted viewpoint images. The multi-baseline stereo is performed by making the matching judgment from an accumulated evaluation function of different image pairs with different baseline. Modification of the original multi-stereo algorithm is carried out to accommodate the fact that the viewpoint image is generated in a different way from a traditional 2D image. Mathematical analysis and the experimental tests performed using the modified multi-baseline algorithm are given. The experiments prove the effectiveness of the modified multi-baseline algorithm described, both on computer generated and captured II.

Improvements are reported using neighbourhood constraint and relaxation technique with a relatively complicated score function rather than the simple SSD function used in the matching task. This is shown to result in a more accurate disparity estimation being obtained regarding the matching position. To improve the computation efficiency, a multi-candidate prescreening technique is also adopted. Experiments show an obvious improvement on the depth map generated from the algorithm, both on the computer generated and captured II.

To deal with the two most difficult situations in matching analysis, untraceable (homogenous) region and untrackable (object border, object occlusion and reappearance) region, "feature block pre-selection" and "consistency post-screening" techniques are used in Chapter 6.  The principle is to remove the false matching results by identifying them either before or after the matching process.  The "feature block pre-selection" is implemented by evaluating the variance within a matching block and the "consistency post-screening" is implemented by evaluating the residue from the score function.  Experiments show the benefits of using the two techniques to rule out invalid disparity results existing in the disparity map.  Acceptable depth maps are obtained for both the computer generated and captured II data, including an image with a natural scene as background.

Finally, Chapter 7 gives a summary of the research undertaken and the achievements made.  Possible further developments of the current research work are also suggested.

# Chapter 2

# Overview of 3D Imaging Techniques

Traditional 2D displays like cathode ray tubes (CRT) or liquid crystal displays (LCD) are popular for a wide range of applications. However, in current applications, these systems display the spatial information from only one perspective view. More and more visual imaging applications need to be able to portray natural and graphically generated environments in three dimensions. The greatly improved sensations of depth and naturalness provided by a 3D display can cause viewers to perceive an increase in the overall picture quality, leading them to prefer 3D presentation (Motoki1995, Taub2002, 3dcgi).

A large variety of 3D imaging systems have been reported by a number of independent research groups, either as general-purpose display units or targeting a specific application (Siegel1995, McCormick1995, Dodgson1997, Outwater1999, Kanghans2002). Among them, Autostereoscopic display provides 3D perception without the need for special glasses or head gear. This chapter gives a brief overview of

3D display technologies, concentrating on Autostereoscopic (free-viewing) systems. An outline of the construction and operation of the integral 3D imaging system developed by the 3D Imaging Group at De Montfort University (DMU) is then presented. Finally, a brief description of the history and current situation of the development of 3D Television is given.

# 2.1 Three dimensional imaging systems

## 2.1.1 3D with glasses (stereoscopy)

3D displays which require the viewer to wear special glasses are reasonably well known. The earliest type of 3D display is stereoscopic imaging which can be traced back to Wheatstone's work in 1832 (Valyus1966). Figure 2.1 shows the Wheatstone stereoscope, in which two geometric 2D drawings exhibiting disparity are viewed using a special device called a reflecting mirror stereoscope. Figure 2.2 is a top view of it. These primitive devices operated by displaying two images side-by-side, one for each eye. Therefore, the two views are channeled separately to the corresponding eye by the simple optical elements.

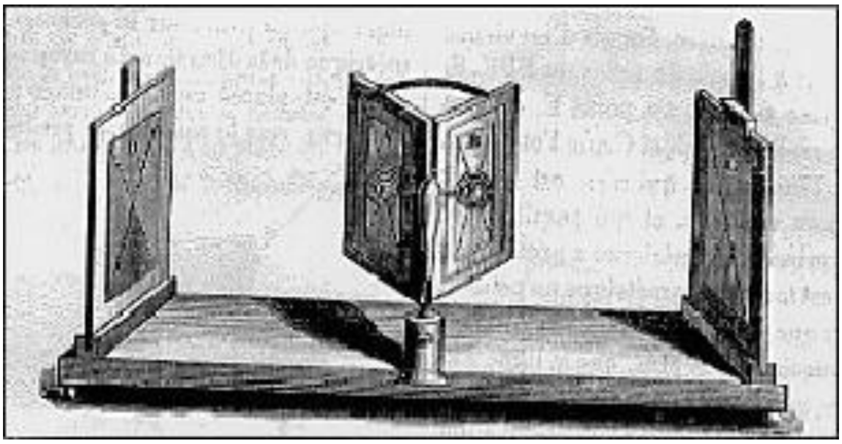


Figure 2.1: Wheatstone's stereoscope ([Okoshi 1976])

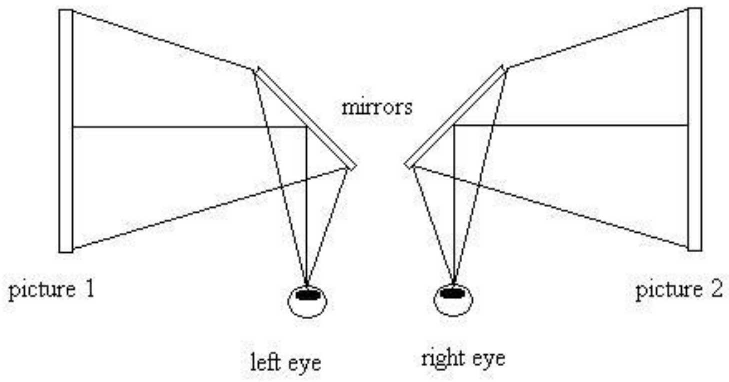


Figure 2.2: Top view of Wheatstone's stereoscope ([Okoshi 1976])

In later developments of stereoscopic displays, viewers are required to wear special image selective glasses. The glasses themselves select which of the two images is visible to each of the viewer's eyes. The technology can divide in three categories (Okoshi1976, Benton1980, Kratomi1972, McAllister1993):

I) Anaglyph method: The left-eye and right-eye images are projected or displayed using two different colours (red/green, cyan/magenta, etc). The viewer observes the images through a pair of colour glasses. The glasses act as a selective device so that each eye only sees one corresponding image.

II) Polarization method: The left-eye and right-eye images are projected or displayed through Polaroid filters and are polarized orthogonally. The viewer observes the images through a pair of Polaroid glasses so that each eye will perceive only one image.

III) Time division method: The two images are sent alternately to the two eyes using a double frame rate display combining with shuttered glasses. The shuttered glasses shield each eye alternatively and present the left-eye image to left eye and the right-eye image to the right eye.

Stereoscopy is one of the simplest ways by far to provide 3D perception using 2D display systems and many entertainment and commercial systems are still using and developing these technologies. The principle exists in using various channeling methods to separate left/right eye images to the corresponding eye. However, the various channeling methods impose their own disadvantages. For example, the anaglyph display, which presents the left and right components of a stereo pair simultaneously using colour coding, produces eye strain due to rivalry between each eye and suffers from colour rendition problems. Polarizing or time division shuttering methods produce similar eye strains as light reaching the eyes is attenuated. Most of all, it is not comfortable for the viewer to wear a pair of glasses for long periods in order to see the 3D effect. This limits the acceptance and applications of the stereoscopic display.

## 2.1.2 Autostereoscopic displays

Autostereoscopic display systems (free-viewing) are more acceptable to observer and therefore are more commercially viable. Auto-stereoscope systems can be subdivided into four types: Create the viewing of an object though reconstruction of the wave fronts, (Holography ); Create viewing that occupies a true volume space (Volumetric Displays); Use complex tracking systems to channel the correct information to the viewer (Two-view Display, Head-tracking Image System and Multi-view Display System); Use a physical sampling optical device to encode the angular information contained in a scene (Integral Imaging System).

### 1) Holography

A hologram is a light wave interference pattern recorded on photographic film (or other suitable surface) that can produce a 3D image when illuminated by suitable light. The principle of holography was established by Denis Gabor (Gabor1948, Gabor1949). Figure 2.3 illustrates how a hologram is recorded (Amateur Holography). A coherent light source, laser beam, is required in the capture process in order to produce interference fringes on the recording medium. The laser beam is split into two beams. The reference beam is spread by a lens or curved mirror and aimed directly at the film plate. The object beam is spread and aimed at the object. The object reflects light and the reflected light is incident on the holographic film-plate. The two beams (reference and reflected) then interact and form a recorded interference pattern on the film. During the replay process, a 3D image of the original object appears when the hologram is illuminated from the original direction of the reference beam. The produced 3D images are virtually indistinguishable from real objects. By shifting position, the viewer can look around or over objects in the foreground to see what is behind them.

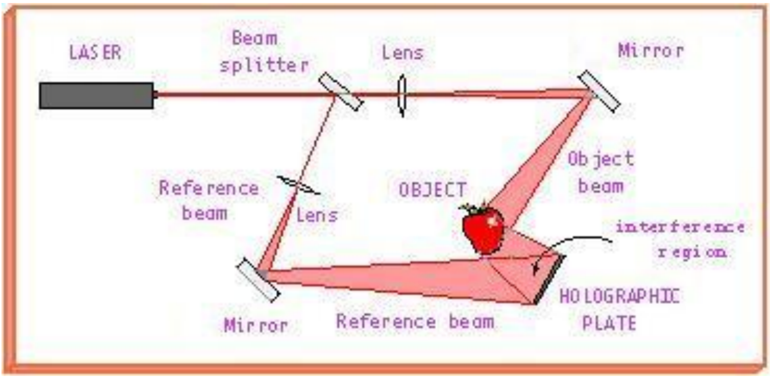


Figure 2.3: An illustration of holography recording ([Amateur Holography])

The unique characteristic of holography is the idea of recording both the phase and the amplitude of the light waves from an object. Since all recording materials respond only to the intensity in the image, it is necessary to convert the phase information into variations of intensity. Holography does this by interfering coherent light from the object and the reference beams derived from the same source.

Many variations on and enhancements to the basic process have been proposed and demonstrated incorporating variations in both viewing conditions and fabrication techniques (Benton1980, Kasper 1987, Outwater1999, Hariharan2002). Depending on the precise materials and technique used in the capture process, reconstruction may take place under coherent or incoherent illumination. The white light reflection hologram, which can be found on credit cards, magazine covers and recording packaging, etc, is perhaps one of the best known developments from the earlier hologram. However, problems exist when attempting to transfer the holographic technique to the display of moving spatial images. An additional problem is the capture of natural scenes as there are the requirements in making hologram for coherent light sources, dark room conditions and high mechanical stability during recording. These considerations reduce the practical utility of the holography technique for general 3D spatial video imaging applications.

## 2) Volumetric displays

Most volumetric displays are portrayed as large transparent spheres with imagery seeming to hover inside (Lewis1971, Kanghans2002). The most common architectures use a rotating projection screen, a gas or a series of liquid crystal panels. As an example, Figure2.4 illustrates a swept-screen volumetric display system. In this system, a three-dimensional data set is first converted into a series of "slices," similar to thin slices of an apple around its core. These slices are stored in a memory bank. A high-speed digital projector illuminates a rotating screen with hundreds of slices of voxel data at a reasonable frequency. Viewers perceive a sharp 3D image due to the fusion caused by the persistence of vision in the eye.

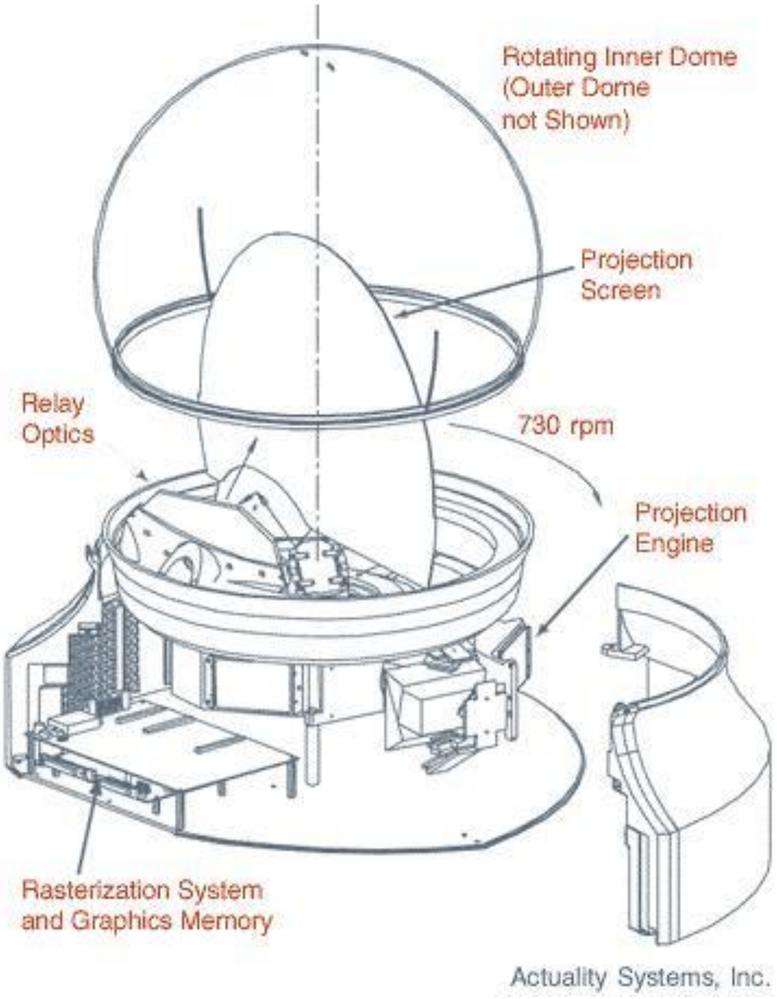


Figure 2.4: A swept-screen volumetric display system ([Actuality Systems])

One benefit of volumetric displays is that they often have large fields of view, such as 360 degrees around the display, viewable simultaneously by an almost unlimited number of people. Volumetric displays are always difficult to design and make, which limited its application in 3D display area.

### 3) Two-view Display, Head-tracking Image System and Multi-view Display System

The simplest and earliest Autostereoscopic imaging technique is the “Parallax barrier method”, which achieves image channeling by interposing a grid-like barrier between the viewer and the display screen (Okoshi 1976). The stereo-pair images are interleaved in alternate strips in precise register with the grid screen so that the left eye can only see the strips from the left eye image and the right eye can only see the strips from the right eye image. This is illustrated in Figure 2.5. The downside of this technique is that only if the viewer stands at the ideal distance and in the correct position can both eyes see the correct images and perceive a stereoscopic effect. In addition, the viewer perceives a reduced brightness since the barrier is used between viewer and the image.

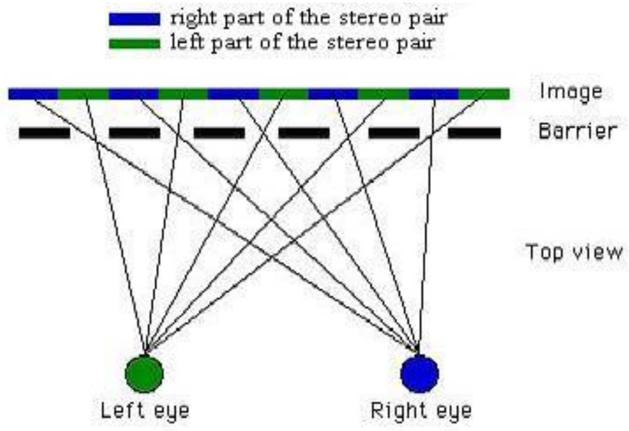


Figure 2.5: Viewing a parallax barrier display ([Bourke1999])

If the position of the viewer’s head is known, the appropriate images (left and right), could be displayed to the appropriate zones adaptively so that each eye can always see the correct images. This is the principle used in the head tracking systems (Tetsutani1994). The limitation of most head tracking systems is that they are single-viewer since the display can only be adjusted to one viewer’s eye position. This is only acceptable in some applications. Beside, it would be pointless to replace the wearing of special glasses with the wearing of a special head tracker for viewers to view the 3D effect.

A similar methodology is used in the lenticular screen method, where the lenticular lens is used as a directional selective scene to separate the left-eye and right-eye view to correct eyes. The simplest form of such a display is comprised of alternating vertical strips of the left-eye/right-eye images, the same as in parallax barrier display, see Figure 2.6. However, each stereo pair of view strips is located precisely behind a lenticular screen during the display rather than the parallax barrier. The lenticular screen works in a similar way to the parallax barrier in directing the corresponding images to different eyes. The advantage over the barrier method is that refraction rather than occlusion is used hence better image brightness can be achieved from the lenticular screen display.

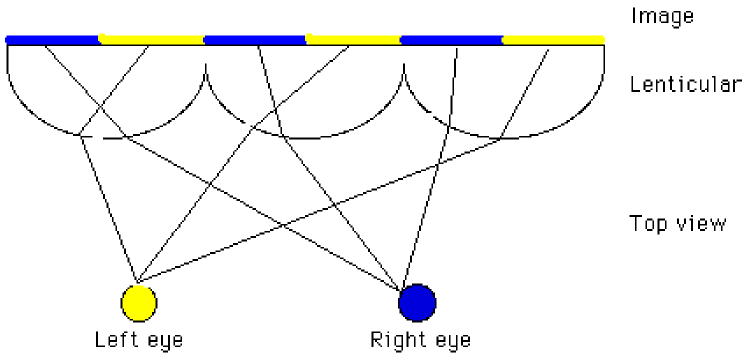


Figure 2.6: Viewing a lenticular screen display ([Bourke1999])

To integrate a “look-around” capability and increase the number of viewing zones, multiple stereo viewpoints are provide by locating bands derived from many views (Harman 1996, Dodgson1997), therefore allowing the viewer to move their head from side to side and see different aspects of the 3D scene. In recent years, many three dimensional television recording and display systems have concentrated on the method of using of multi-camera capture and multi-view displays to allow a certain degree of look-around capability. The viewing effect depends on the number of viewpoints. The difficulty of simultaneously generating or capturing enough views in real time at an affordable cost becomes the major problem in this type of system.

## 4) Integral imaging

Integral imaging is a technique that is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution by using unique optical components. It is akin to holography in that 3D information is recorded on a 2D medium and can be replayed as a full 3D optical model. However, in contrast to holography, coherent light sources are not required for integral imaging. This conveniently allows more conventional live capture and display procedures to be adopted.

All integral imaging can be traced from the work of Gabriel Lippmann (Lippmann 1908), where a micro-lens sheet was used to record the optical model of an object scene. The micro-lens sheet was made up of many micro-lenses having the same parameters and the same focal plane. In the arrangement, the recording film is placed at the focal plane of the micro-lens sheet, as illustrate in Figure 2.7. Following the film development, a full natural colour scene with continuous parallax can be replayed using another micro-lens sheet with appropriate parameters. The replayed image is spatially inverted as shown in Figure2.8

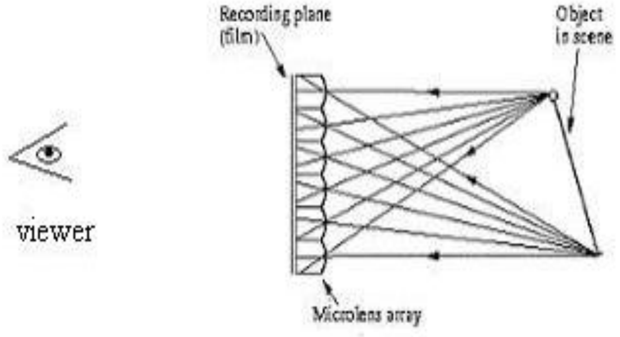


Figure 2.7: The recording of an integral image

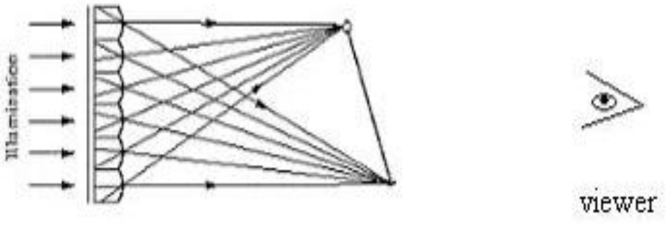


Figure 2.8: The replay of an integral image (The viewer perceives a spatial inverted 3D scene)

To overcome the problem imposed by the pseudoscopic (spatially inverted) nature of the integral image, a modification to the Lippmann system was proposed by Ives (Ives1931), in which a second recording process is introduced before replaying, as shown in Figure 2.9. When the second-stage photograph is replayed, a 3D image with correct spatial depth (orthoscopic) can be observed, see Figure 2.10.

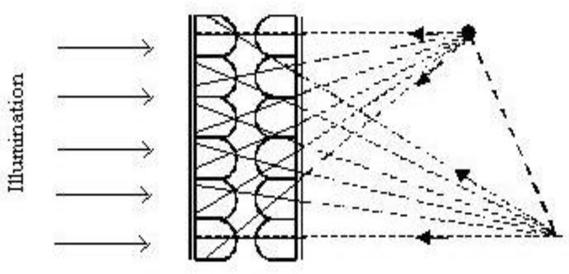


Figure2.9: A second stage recording of integral photograph

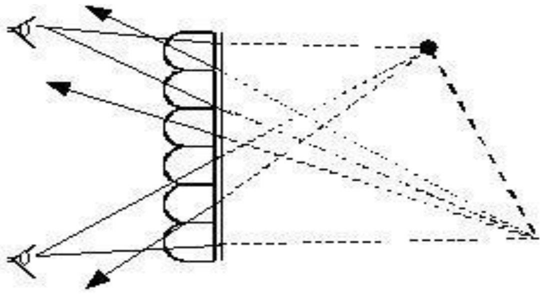


Figure 2.10: Replay and viewing of the orthoscopic image scene

The two-stage recording process can produce an orthoscopic 3D scene with corrected spatial position. However, substantial image quality degradation is introduced due to the distortions introduced by the micro-lenses and film emulsion, stray light, etc. To overcome this problem, a two-tier network as a combination of macro-lens arrays and micro-lens arrays was reported by Davies and McCormick in DMU (Davies1994). The two-tier network works as an optical "transmission inversion screen" which overcomes the image degradation caused by the two-stage recording process and allows direct spatially correct 3-D image capture for orthoscopic replay. Theoretically, this network is able to capture object space from 0.3m to infinity. In consequence, the integral photographic technique pioneered by Lippmann has been improved. With recent progress in micro-lens manufacturing techniques, integral imaging is becoming practical and prospective 3D display technology and hence is attracting much interest.

## 2.2 An advanced integral imaging system

The optics of an advanced form of integral imaging system employing a two-tier optical network was developed and has been described in detail by Davies and McCormick (Davies1988, 1994, McCormick1992, 1994, 1995). The optical arrangement, shown in figure 2.11, comprises two macro-lens arrays placed equidistantly behind and in front of an auto-collimating transmission screen (ATS). The ATS is made up of two microlens arrays separated by their joint focal distance. The recording plane is a photographic plate whose position coincides with the focal plane of another microlens array.

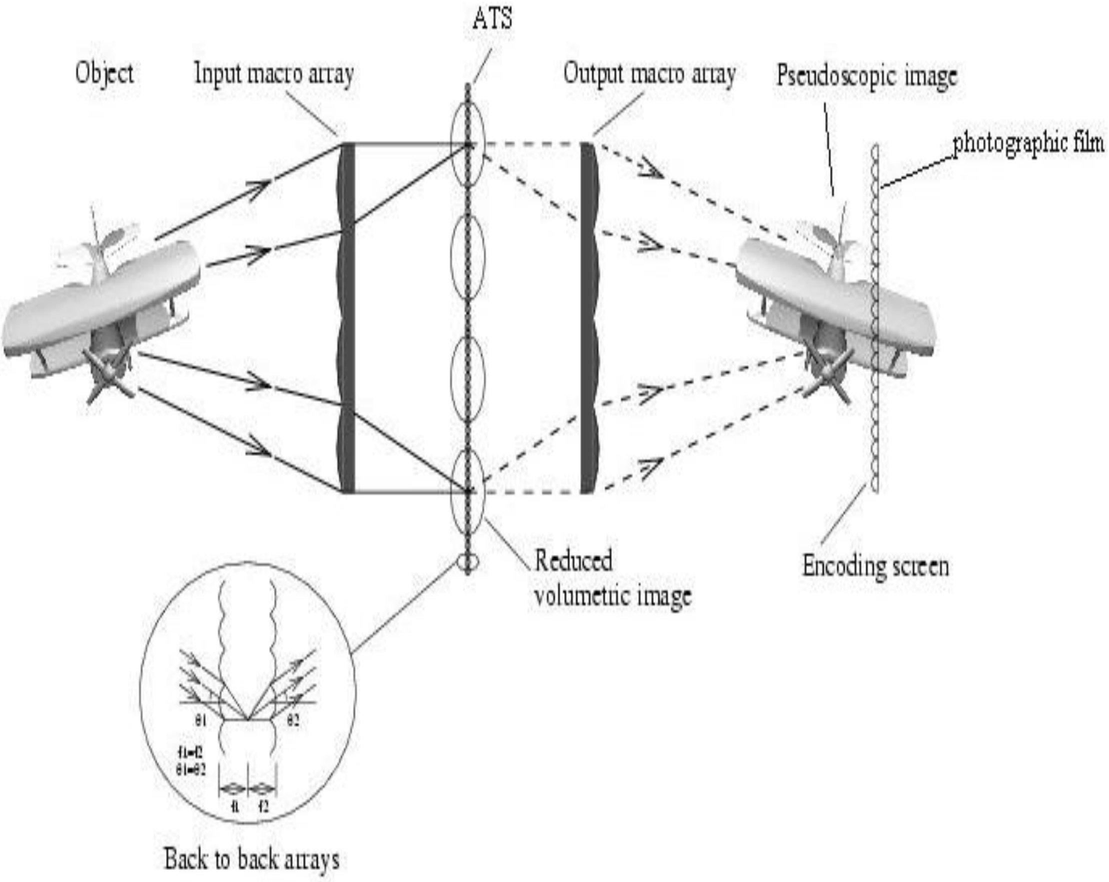


Figure 2.11: The advanced integral imaging system ([Davies 1994])

The optical transmission process of the advanced optical system is illustrated in Figure 2.12. The input macro-lens array first transmits the compressed object space to or near the central double micro-lens screen (ATS). The screen inverts the spatial sense of each intermediary image and simultaneously presents these spatially reversed 3D optical models to the corresponding output macro-lenses. The output macro-lenses array then re-transposes the optical model to the correct spatial location. The final integrated optical model before recording, formed by the second macro-lens array, is a true 3D optical 1:1 reconstruction of the original object.

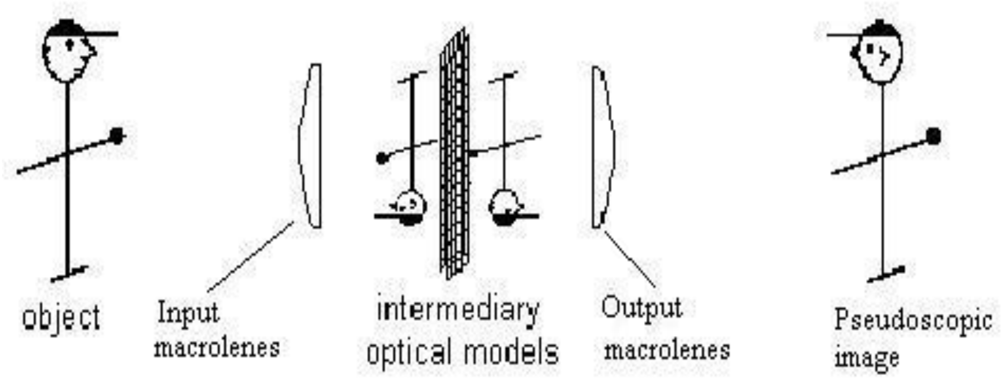
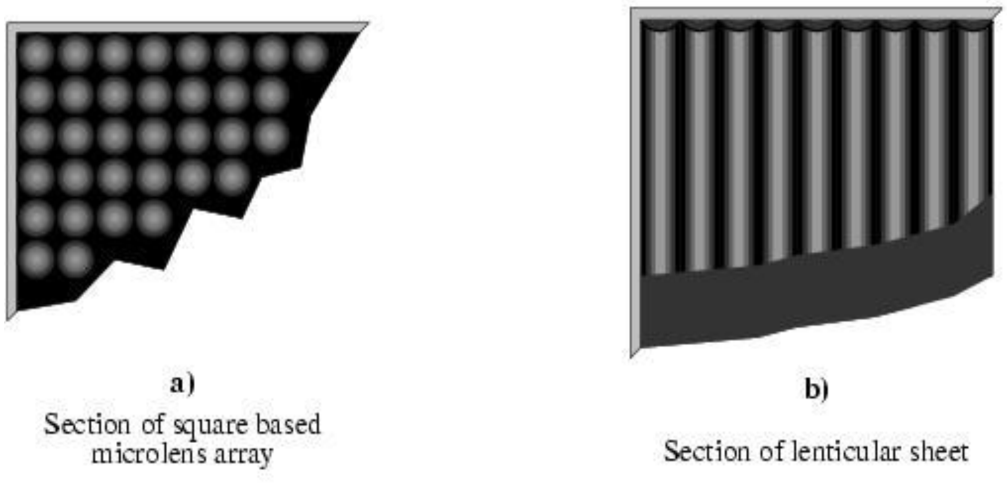


Figure 2.12: The optical transmission within the advanced integral imaging system ([Manolache1999])

The II is recorded on a film using a microlens array put after the two-tier optical network. Each microlens of the recording array samples a fractional part of the pseudoscopic scene, many microlenses record directional information of the scene from different viewing angles. Therefore, parallax information for any particular point is spread over the recording plane. The angular information is further recorded by a film placed at the focal plane of the microlens array. This recorded II can be replayed by overlaying it with a microlens array having the same parameters.

In the above outlined capture and display processes, microlens arrays are used in both the encoding and decoding of the planar intensity distribution. The arrays used for this purpose typically comprise square based spherical lens-lets, which are capable of encoding the object scene with continuous parallax in all directions. A section of such a lens array is illustrated in Figure 2.13a. It is also possible to record and replay integral

3D images using lenticular sheets, which comprise many thin cylindrical lenses (lenticular sheet), as shown in Figure 2.13b. Integral images recorded in this way possess parallax only in one direction. It is worth mentioning that the integral image produced by the lenticular sheet is not the same as multi-view image with lenticular screen display. In the multi-view display system, the lenticular sheet is used merely for spatial de-multiplexing of multiple separate views of an object scene. To distinguish this difference, the term unidirectional integral image (UII) is used in the thesis to represent the image generated by integral imaging technique using lenticular sheet. The term Omni-directional integral image (OII) is consequently used to represent the integral image formed using the square based spherical microlens arrangements in recording.



a) Section of square based microlens array

b) Section of lenticular sheet

Figure 2.13 Diagrammatic representation of the lens array ([Forman 1999])

## 2.3 Three dimensional Television (3D TV)

Among the many applications of 3D image technology, 3D TV attracts the strong interest of the entertainment industries. However, the advent of 3D TV has been slower than many predicted. "Three-dimensional television is like the circus: it comes and it goes, and people line up to see it every time it appears," remarked Daniel Symmes, president of Dimension 3, a 3-D Television production company in Woodland Hills, Calif. “3D TV is a bigger change to television than the switch from standard definition to HDTV", states Chris Yewdall, president of Dynamic Digital Depth (DDD), etc. The question, now as then, is whether 3-D has staying power or will remain a gimmicky fad? (Taub2002)

Almost everyone has seen or heard of 3D television shows or 3D theme park attractions. It is recognized that an acceptable 3D TV system should ideally be:

I) Autostereoscopic (does not require glasses),

II) Transmittable within existing broadcast bandwidth, full natural colour suitable high resolution flicker free displays.

III) Compatible with 2D such that 3D signals can be accepted and displayed by a 2D receiver as a 2D image.

Most of the 3D TV developed in the past has relied upon the use of special glasses to experience the three dimensional effect. Today, many companies are aimed at developing technologies that can remove the need for glasses. These include Actuality System's volumetric display (Perspecta), Deep Video Imaging's 3D display (actualdepth), Dimension Technologies's Virtual Window™, The Dresden 3D GmbH (D4D), etc. Most of the techniques work on the principle of presenting multiple images to the viewer by use of temporal or spatial multiplexing of several discrete viewpoints to the eyes. Recently, the DDD company together with screen manufacturer, 4D-vision, claimed that 3D TV without glasses has become a reality. In the developed system, the depth information is stored in a small separate data channel that can be transmitted along with a broadcast or DVD signal and decoded using a special set-top box. For a good viewing effect, a great number of planes are needed and it takes a long time to create the proper look for just one still image. Therefore, the problem of simultaneously generating or capturing enough views in real time at an affordable cost still exists.

A schematic of 3D TV based on the integral imaging technique is illustrated in Figure 2.14 (McCormick1995). As mentioned in the previous section, a pseudoscopic optical model of the object scene is produced by the two-tier transmission unit in one recording. The encoding microlens array produces an intensity distribution of the object scene at its back face. At this point, instead of placing a photographic film, a copy lens is used to copy the current planar encoding of the object scene on to a high resolution CCD imaging device. The recorded 3D data is then transmitted to the user after coding. Both flat panel and projection displays can be used as output devices. Unlike multi-view systems, the advanced integral imaging system has the capability of capturing the 3D scene in one single stage with continuous parallax, providing good live display practicality and look-around capability in real time. Despite the fact that there are a number of technological challenges that need to be overcome, the 3D integral imaging technique probably has the most potential in producing future 3D TV.

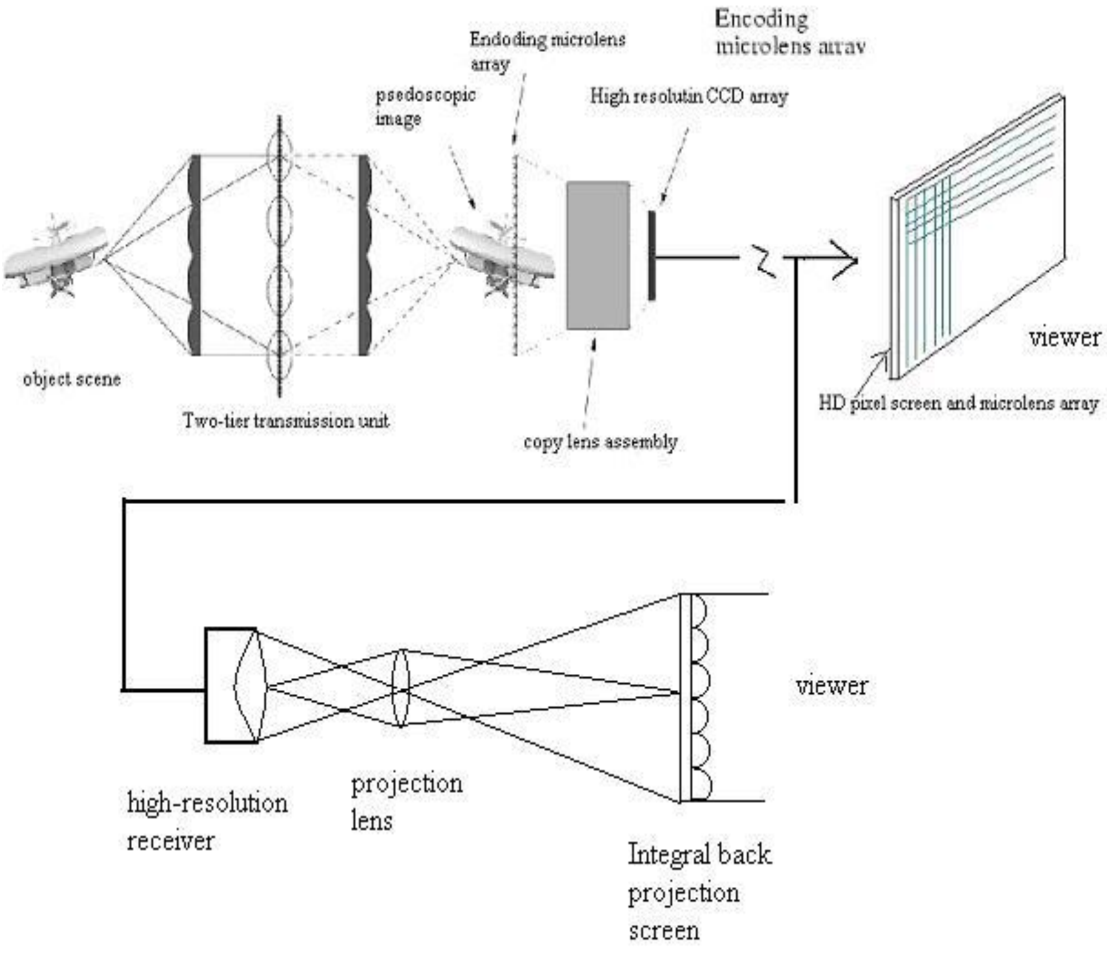


Figure 2.14: A schematic of 3D TV ([McCormick 1995])

## 2.4 Summary

This chapter gives a brief review of 3D imaging techniques, concentrating on free-view (Autostereoscopic) systems, followed by a detailed description of the 3D integral imaging system developed by the 3D Imaging Group at De Montfort University. As a promising technique in 3D display, the application of integral imaging in 3D TV has been envisaged.

# Chapter 3

# Review of the depth extraction methods from 3D images

Depth information is used in many applications, for example, 3D modeling of natural objects, 3D remote handling and quality control, virtual studio and 3D telepresence. In chapter 2, the 3D imaging system at De Montfort University is described by which a 3D image of true optical models can be recorded. The replayed image demonstrates continuous parallax in all directions. If successfully engineered, this 3D imaging system could have enormous implications for both leisure and industrial applications. One of the major potential applications is 3D TV systems. In this application, the generation of a depth map is essential if real and/or computer generated objects are to be integrated within integral 3D TV images. It is also essential to enable content-based image coding and content-based interactive manipulation to be carried out.

The simplest and most convenient way of representing and storing the depth measurements taken from a scene is a depth map. A depth map is a 2D array where the x and y distance information corresponds to the rows and columns of the array as in an ordinary image, and the corresponding depth readings (z values) are stored in the array’s elements. It is similar to a grey scale image but the depth information is used to replace the intensity information. The aim of the present work is to generate a depth map from 3D integral image (II) data. It is therefore useful to both consider the techniques used for depth estimation in current computer vision applications and to acquire a deep understanding of the 3D II data.

## 3.1 Common approaches to obtain depth information

### 1) Stereo vision

Perhaps the most common approach in obtaining information on the 3D structure and distance within a scene is stereo vision (Forstner 1986, Hannah 1989, Matthies 1989, Trucco 1998, Okutomi 1993). Figure3.1 illustrates a top view of a stereo system composed of two pinhole cameras. The left and right image planes are coplanar, and are represented by the segments $I_l$ and $I_r$ respectively. $O_l$ and $O_r$ are the centers of projection. The optical axes of the two cameras are parallel.

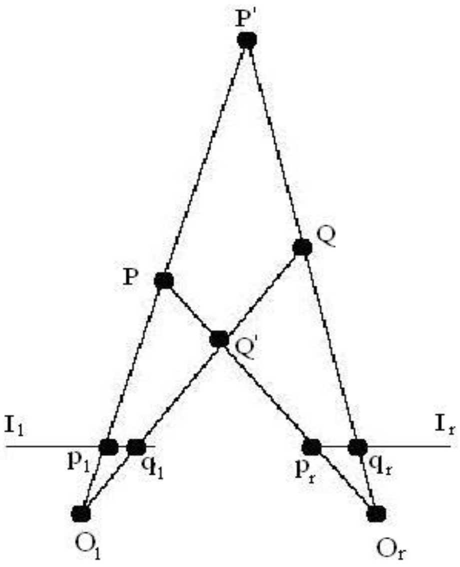


Figure 3.1: A simplified stereo imaging system

The method by which stereo determines the position in space of P and Q is triangulation, that is, by intersecting the rays defined by the centers of projection and the images of P and Q, $p_l$ , $p_r$ , $q_l$ , $q_r$ . Triangulation depends crucially on the solution of the correspondence problem: If ($p_l$ , $p_r$ ) and ($q_l$ , $q_r$ ) are chosen as pairs of corresponding image points, intersecting the rays $O_l p_l - O_r p_r$ and $O_l q_l - O_r q_r$ leads to interpreting the image points as projections of P and Q. However, if ($p_l$ , $q_r$ ) and ($q_l$ , $p_r$ ) are the selected pairs of corresponding points, triangulation will return the wrong object point $P'$ and $Q'$.

Provided that the correspondence problem has been solved, the position of a single point P or Q can be recovered from the corresponding projections on $I_l$ and $I_r$. Figure 3.2 illustrates the estimation of depth from the corresponding projection points in two images. Consider the single object point, P, from its projections, $p_l$, $p_r$. The distance ($\Delta$) between the two cameras project center, is called the baseline in a stereo system. If $x_l$ and $x_r$ are the coordinates of $p_l$ and $p_r$ with respect to the principal points $c_l$ and $c_r$, f is the common focal length of two cameras and Z is the distance between P and the baseline, from the similar triangles ($p_l$ , P, $p_r$) and ($O_l$ , P, $O_r$), we have:

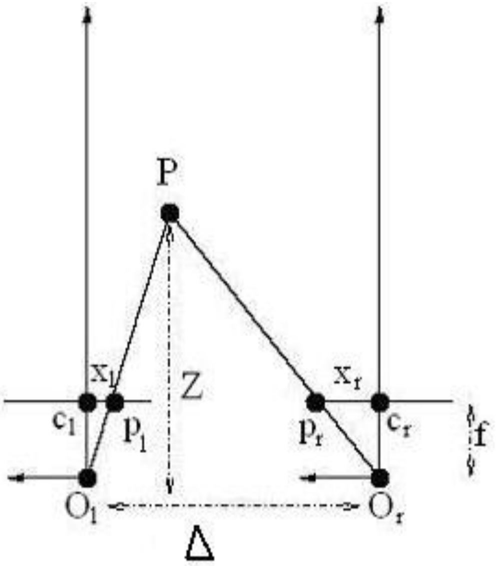


Figure 3.2: depth estimation in stereo vision

$$\frac{\Delta + x_l - x_r}{Z - f} = \frac{\Delta}{Z} \qquad (3.1)$$

Solving (3.1) for the depth Z, gives:

$$Z = f\,\frac{\Delta}{d} \qquad (3.2)$$

where, $d = x_r - x_l$ is the disparity, which is the difference between the corresponding points in the two images. Here, the depth of a point (*Z*) is inversely proportional to the disparity (*d*). A big disparity means the object point is close to the recording camera. Normally, $\Delta$ and *f* are fixed in stereo vision, the depth of the object point can be determined from the disparity.

On the other hand, for an object point at a particular depth, the disparity is proportional to the baseline ($\Delta$). This implies that if there is a fixed error in determining the disparity then the accuracy of depth determination will increase with $\Delta$ increase. Since the disparity can only be measured in pixel differences, increased $\Delta$ will lead to better accuracy in depth measurement. However, as the camera separation becomes large, difficulties arise in finding the corresponding points in the two camera images. In order to measure the depth of a point it must be visible to both cameras and we must also be able to identify this point in both images. However, as the camera separation increases, the differences in the scene recorded by each camera will increase. It becomes increasingly difficult to match corresponding points in the images. This is known as the stereo correspondence problem. Corresponence problem is the primary computational problem for stereo vision. Different approaches have been researched concerning this problem(Mori 1973, Lucas 1981, Forstner 1986, Barnard 1989, Okutomi 1993, Trucco 1998).

### 2) Structure from motion

An alternative approach for obtaining information on the 3D structure within a scene is to extract depth information from the spatial and temporal changes occurring in an image sequence. In most literature, this is referred to as the method of obtaining structure from motion (SFM). The SFM can be described as: Given a series of images taken at different times with moving objects in the scene or a scene filmed by a moving camera, useful information about the scene can be obtained by analysing and understanding the difference between images caused by the motion. Over the past decades, SFM has been a central problem in computer vision. The literature contains a variety of schemes for dealing with this issue (Aggarwal 1988, Huang 1994, Srinivasan 2000, Dallaert 2002). Among them, approaches to calculating optical flow and setting up relationships between optical flow and 3D structure/motion parameters are the most popular particularly as no prior knowledge is required about the point of correspondence.

The basic idea of SFM from optical flow involves two steps:

1) The use of the **image brightness constancy equation** to calculate the components of the motion field (Optical flow).

2) Using the **basic equation of the motion field** to calculate the depth.

The image brightness constancy equation is described as:

$$(\nabla E)^T v + E_t = 0 \qquad 3.3)$$

Where, E (the image brightness) is a function of the spatial coordinates of the image plane, *x,y* and time *t*. E=E(*x,y,t*). The motion field *v*, has two components, $v_x$ and $v_y$. In components, the image brightness constancy equation can be written as:

$$E_x \cdot v_x + E_y \cdot v_y + E_t = 0 \qquad 3.4)$$

The basic equation of the motion field is described as:

$$v = f \frac{ZV - V_z P}{Z^2} \qquad 3.5)$$

Where, P=[X,Y,Z]$^T$ is a 3D point in the usual camera reference frame and $V = -T - w \times P$ is the relative motion between P and the camera. In components, the relationship can be written as:

$$v_x = \frac{T_z x - T_x f}{Z} - w_y f + w_z y + \frac{w_x xy}{f} - \frac{w_y x^2}{f} \qquad 3.6)$$

$$v_y = \frac{T_z y - T_y f}{Z} - w_x f + w_z x + \frac{w_y xy}{f} - \frac{w_x y^2}{f} \qquad 3.7)$$

The SFM through calculating optical flow method involves complicated mathematical computation. For simplicity, a restricted case where the camera is known to be translating along the X (horizontal axis), as shown in Figure 3.3, is given as an example. In the example, the camera has focal length f which is oriented along the Z axis and moves only along the X direction.

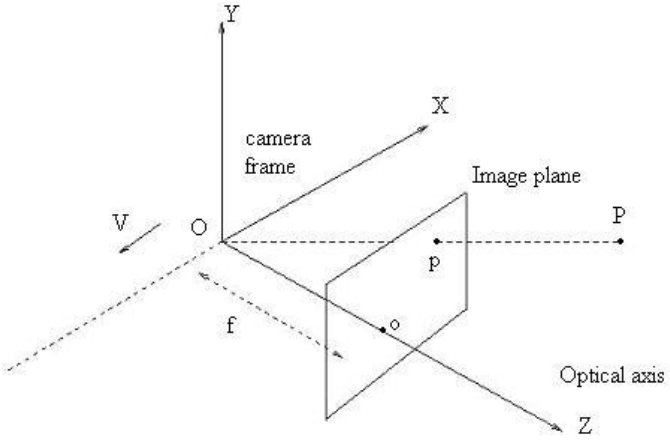


Figure3.3: The camera system used in the example

Under the assumption of the restricted camera moving in the example, the basic equations of the motion field 3.6) and 3.7) can be simplified as:

$$v_x = -\frac{f}{Z} T_x \qquad 3.8)$$

$$v_y = 0 \qquad 3.9)$$

Substituting the motion field equation into the image brightness constancy equation, gives:

$$E_x \cdot (-\frac{f}{Z} T_x) + E_y \cdot 0 + E_t = 0 \qquad 3.10)$$

Therefore, the depth Z of a point can be easily derived out from equation 3.10) in the example:

$$Z = -\frac{E_t}{E_x} fT_x \qquad 3.11)$$

Where, $E_t$, $E_x$ can be estimated directly from the image or the image sequence.

As only one camera is needed to recover the depth information and a depth solution can be found without correspondence, the SFM based on optical flow method has gained wide interest by researchers in last two decades. However, it is worth noticing that the validity of the image brightness constancy equation is only under the assumptions of Lambertian surfaces, pointwise light source is at infinity and there is no photometric distortion. In addition, the optical flow obtained from the equation is only an approximation of the normal component of the motion field. It often performs poorly in highly textured and fast motion regions. Besides, the method is very sensitive to noise in a practical case since derivation is involved in the calculation.

## 3.2 Depth extraction from integral images

### 1) Depth from disparity

The mathematical model of the integral imaging system developed in 3D image group has been analyzed and described in detail by Manolache (Davies1994, Manolache1999). Figure 3.4 illustrates the two tier optical network for a UII camera system with associated Cartesian coordinate system Oxyz. The Z-axis denotes the depth direction, while the x, y-axes describe the lateral positions.

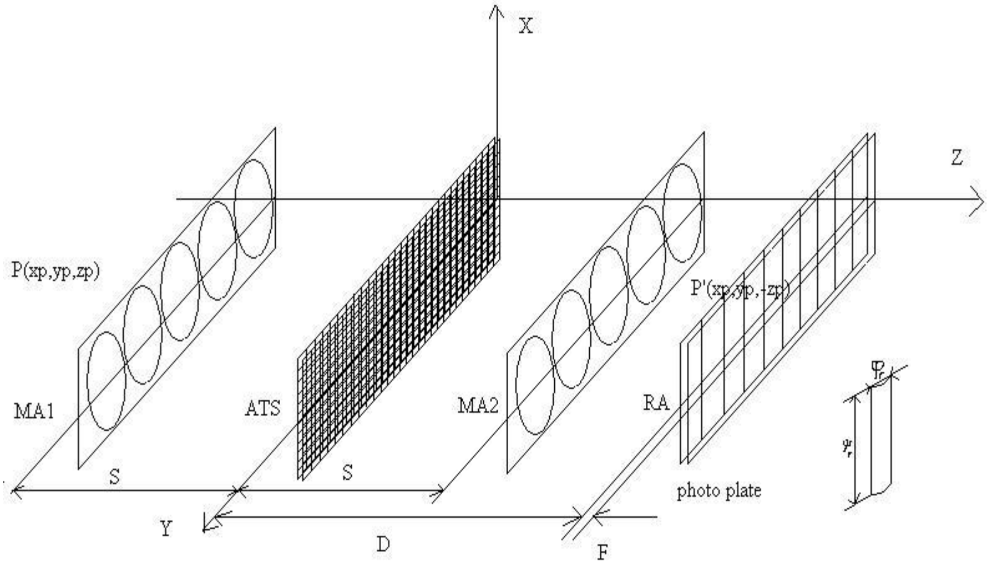


Figure3.4: 3D UII camera system using a two tier optical network ([Manolache 1999]).

It is proved that as a result of the optical process of the two macrolens array and the ATS screen, each object point P($x_p$,$y_p$,$z_p$) is reconstructed as an optical model formed by intersecting modulated ray bundles at the location P'($x_p$,$y_p$,-$z_p$), which is the equal conjugate image location of point P with respect to the Oxy plane (Manolache1999). In the integral recording, the intensity distributions related to the integral image P' of P are recorded on a photographic plate that lies behind the recording microlens array. Each microlens of the recording array records a fractional part of the scene, many microlenses recording directional information of the point from different viewing angles. Therefore, parallax information about this particular point is spread over the recording plane, as shown in Figure 3.5. The recorded data under *k*th microlens, centred at the point $P_k$', has coordinates:

$$x_k^{'} = x_p \quad 3.12)$$

$$y_k^{'} = \frac{(c_l + \varphi_r k)(D + F - |z_p|) - y_p F}{D - |z_p|} \qquad 3.13)$$

$$z_k^{'} = D + F \qquad 3.14)$$

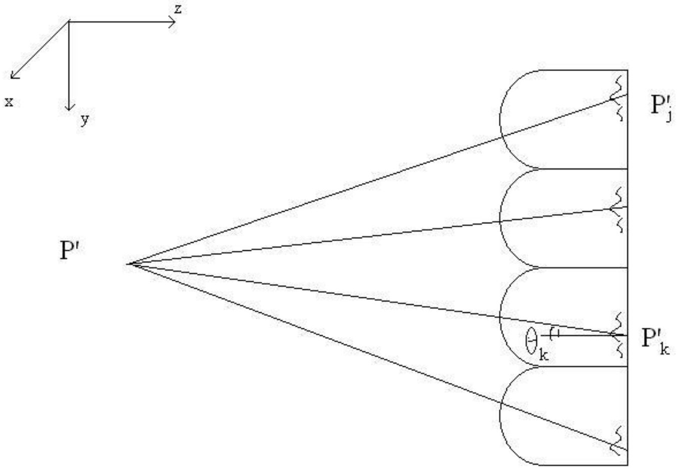


Figure 3.5: microlens recording

Where, $c_l$ is the coordinate of the first microlens centre.

Similar equations exist for point $P_j$’. Thus, the disparity between two recorded intensity distributions of the point P corresponding to microlenses k and j can be written as:

$$d_{jk} = \frac{(D + F - |Z_P|)\,|\,j - k\,|\,\phi_r}{D - |Z_P|} \qquad 3.15)$$

This expression allows the recovery of the position of a physical point when the disparity between the recording, centred at $P_j^{'}(x_j^{'}, y_j^{'}, D + F)$ and $P_k^{'}(x_k^{'}, y_k^{'}, D + F)$ are known. Namely:

$$x_p = x_j^{'} = x_k^{'} \qquad 3.16)$$

$$y_p = \frac{(c_1 + \phi_r j) d_{jk} - y_j^{'}\,|\,j - k\,|\,\phi_r}{d_{jk} - |\,j - k\,|\,\phi_r} \qquad 3.17)$$

$$z_p = \frac{(D+F)\,|\,j-k\,|\,\phi_r - d_{jk} D}{d_{jk} - |\,j-k\,|\,\phi_r} \qquad 3.18)$$

The equation 3.18) allows retrieve depth from the disparity mathematically.

However, in a practical case, each microlens used in the integral recording can be regarded as a very small low resolution camera. In a typical case, using a 600um pitch sized lenticular sheet in recording, the pixel numbers corresponding to the small pitch size is no more than 30 pixels. Therefore, the problem of matching corresponding intensity distributions under each microlens is very difficult due to the very low resolution achievable. Consequently, deriving the depth of a point using equation 3.18 directly from the disparity is practically impossible.

## 2) Point spread function in the 3D unidirectional integral recording system

In the previous work, the role of diffraction due to the very fine pitch of a microlens involved in this optical process is further considered. It is considered that each object point will give rise to an intensity distribution in the image, shown in Figure 3.5. This intensity distribution function is called the point spread function (psf). The calculation of the point spread function of the II system has to take into account the three microlens arrays (ATS screen plus the recording microlens array) through which light pass on. The point spread function behind a specific microlens *k* of the recording array of point P can be obtained by using the Fresnel-Kirchhoff formula (Manolache1999):

$$psft_k(x,y) = \frac{v^4 w_1 w_2}{\sqrt{(v^2+w_1^2)(v^2+w_2^2)}} \exp(-\frac{(x-x_k^{'})^2}{v^2+w_1^2}) \exp(-\frac{(y-y_k^{'})^2}{v^2+w_2^2}) \qquad 3.19)$$

With an inclination factor correction, the point spread function of the whole process is given as (Manolache1999):

$$psf_{total}(x,y)=\sum_{k}\cos^{2}\theta_{k}\, psft_{k}(x,y) \qquad 3.20)$$

where, $\theta_k$ is the angle subtended to the normal by the recording ray P'P'$_k$.

$$w_1^2=0.106\frac{\psi_r^2 b^2 F^2}{a_r^4}+0.245\frac{\lambda^2 F^2}{\psi_r^2} \qquad 3.21)$$

$$w_2^2=0.106\frac{\varphi_r^2 b^2 F^2}{a_r^4}+0.245\frac{\lambda^2 F^2}{\varphi_r^2} \qquad 3.22)$$

$a_r$ is the distance between the reconstructed model P' and the recording microlens array ($a_r = D-|Z_P|$), and $\psi_r$, $\varphi_r$ and F are the length, width and the focal distance of a recording microlens, $\lambda$ is the wavelength.

From above equations, it can be seen that the shape of the $psf_{total}$ , is a two variable Gaussian function, and the energy forming the recorded intensity distribution is concentrated on a rectangular spot whose dimensions depends on point depth. Therefore, it is reasonable to think of using it as a tool for extracting depth information from the integral images.

## 3) Depth extraction as an inverse problem

When the object space is imaged and recorded, the resulting intensity distribution $I_{rec}$ in the recorded II is the convolution between the point spread function and the object space $I_{obj}$(Manolache2000):

$$I_{rec} = psf * I_{obj} \qquad 3.23)$$

On the other hand, provided the inverse function of *psf*, denoted by *Inv(psf)*, is known, the intensity distribution of the object can be calculated as:

$$\hat{I}_{obj} = Inv(psf) * I_{rec} \qquad 3.24)$$

This computation in equation 3.24) allows the reconstruction of the object space from the recorded image.

In order to provide the numerical model for the above approach, a matrix formulation is necessary. The recorded image has a natural discrete structure. However, the object space is a continuous space and therefore needs to be transformed to a discrete space (a set of discrete points). For the recorded image, assuming the horizontal section of an image is taken with a semi-cylindrical microlens array which contains *m* microlens, and there are *p* pixels on the image under each microlens, the recorded intensity distribution can be represented by a vector with *n=m*p* elements, as shown in figure3.6. If the object space is conceived as a set of points *q* with well determined spatial coordinates, one can associate a *n*q* matrix PSF to the point spread function whose elements PSF[l,k] are the maximum intensity produced by the object point *k* in the image pixel *l*. Under this sampling scheme, the convolutions in formulae 3.23), 3.24) become ordinary matrix multiplications:

$$I_{rec} = PSF \cdot I_{obj} \qquad 3.25)$$

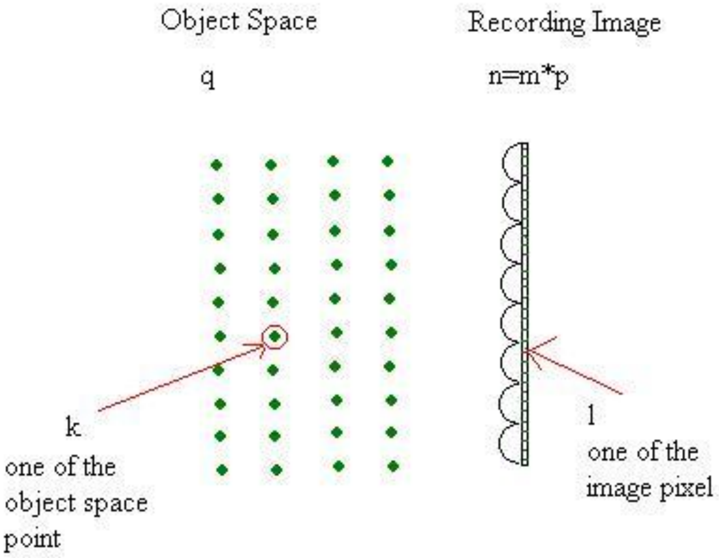


Figure 3.6: Conversion of the continuous space to discrete space

The inverse transformation corresponding to matrix PSF is represented either by the inverse matrix $PSF^{-1}$, if PSF is invertible, or by the Penrose-Moore pseudo-inverse matrix $PSF^{+}$ ( $PSF^{+} = (PSF^{T} * PSF)^{-1} * PSF^{T}$ ).

$$\hat{I}_{obj} = PSF^{-1} \cdot I_{rec} \quad \text{or} \quad \hat{I}_{obj} = PSF^{+} \cdot I_{rec} \qquad 3.26)$$

Each element of matrix PSF gives the corresponding relationship between object point '*k*' and image pixel '*l*'. After reversal, each element of the pseudo inverse matrix $PSF^{+}$[k, l] gives a description of the corresponding relationship between image pixel '*l*' and object point '*k*'.

As an example, assume object space contains a square shaped object and a fence-like background, as shown in Figure 3.7. For simplicity, only one object plane is considered here. A rectangle net (21*15) is chosen for describe the object space, as shown in Figure 3.8. The parameters of the assumed camera system are:

*Camera aperture is 195mm*

*D, the distance from ATS screen to recording microlens array is 341.8575mm*

*$Z_{ref}$, the reference plane of object space is 327.0878mm*

*F, φ, the focal length and pitch size of the microlens sheet is 3.23mm, 1.124mm respectively.*

*$Z_d$,$Y_d$, the sample distance in Z,Y space for object space is 1mm and φ/2 respectively.*

*There are 21 microlenses in recording microlens array and 15 pixels under each microlens.*

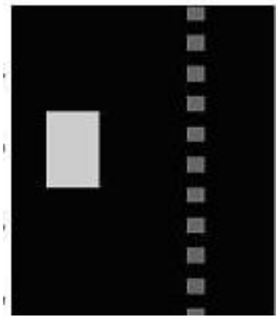


Figure 3.7: An assumed object

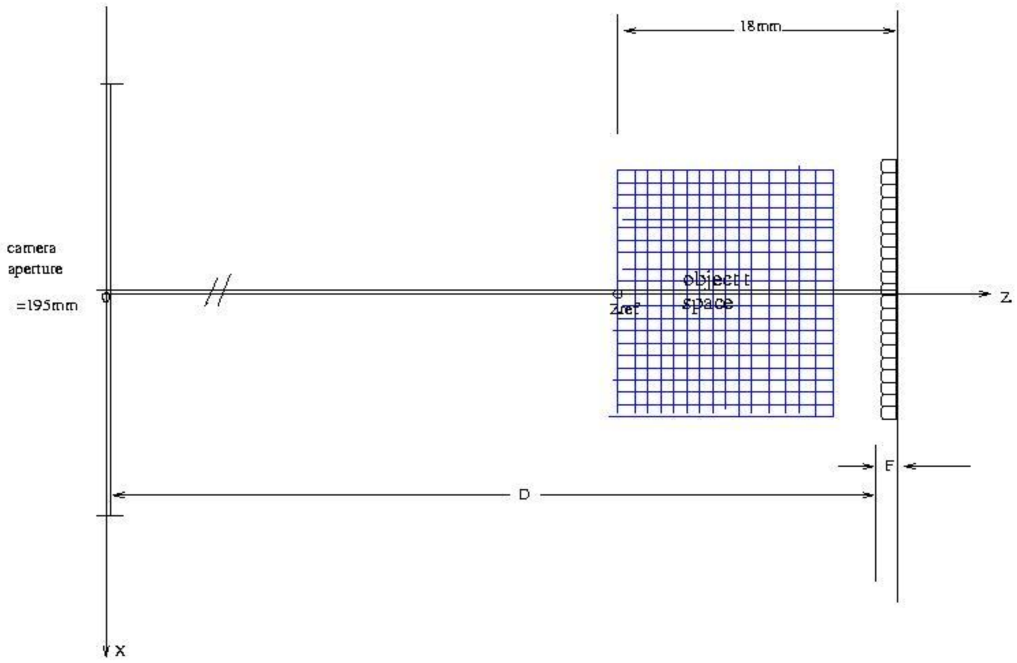


Figure 3.8:  The rectangular net model of the object space.

Under the condition that each small rectangle space (voxel) can be represented as a point on its centre position, a $PSF_{315*315}$ matrix corresponding to the assumed camera system with the defined discrete space can be formed from equation 3.20.  Using the PSF matrix, an UII can be generated on the assumed object shown in Figure 3.6.   The generated UII data is shown in Figure 3.9 which can be replayed by using a microlens sheet with suitable parameters.

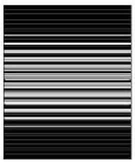


Figure 3.9:  The UII data generated by a PSF matrix from an assumed object

Using the corresponding $PSF^+$ matrix, the object space can be further reconstructed from Figure 3.10.  It can be seen from the reconstructed object space on Figure 3.10 that the original object space has been reconstructed with the correct position except for some small intensity change within the object space in the simulation.

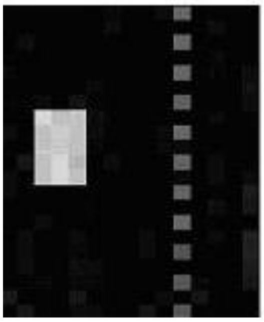


Figure 3.10: The reconstructed object space from the UII data generated by the PSF matrix.

Figure 3.11 is the reconstructed object space from figure 3.8 using an $MPSF^+$ matrix, where all elements of the right half of the PSF matrix are set to zero. This removes all the objects in the rear part of the object space (the fence-like background).

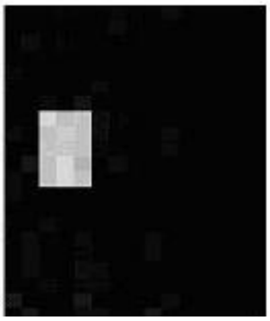


Figure 3.11: The reconstructed object with the fence-like background removed

The above experiment shows the effectiveness of the depth extraction approach based on image inverse theory with a corresponding PSF matrix on simulation data. However, the effort of applying the approach to a realistic integral image proves in vain, both on computer generated or captured images. This is due to the very ill-posed discrete correspondents associated with the direct process

(Bertero1998, Manolache2002). As an improvement, a hierarchical adaptive regularization method is further used in order to obtain high resolution in object space reconstruction and a constrained least squares solution of the depth extraction problem(Manolache2002). The new algorithm proved to be capable of producing high resolution reconstruction and computation efficient in simulation but still could not work out the depth from a realistic II. The question raised is: “When the discrete points are used to represent the object space, what is the maximum size of the voxel that can be represented by one single point in the optical recording process?” To work out the depth extraction task from realistic II’s, an alternative way needs to be considered. The new approach is described in the following chapters in detail.

## 3.3 Summary

This chapter begins with the common depth extraction method used in the computer vision area and then moves to the task of extracting depth information from 3D II’s. Due to the unique recording process existing in the information distribution of the integral image, common approaches used in stereo vision can not be directly applied to obtain depth information from the II. Previous work has attempted to tackle this task as an image inverse problem where the object space is taken as a set of discrete points and a corresponding PSF matrix is used to represent the relationship between object point and the recorded II. The approach achieves accurate results for simulation data but has difficulty when applied to realistic II data due to the inherent loss of the information associated with the discrete model. To produce a workable technique for the depth extraction task applicable to realistic II’s, an alternative way needs to be considered.

# Chapter 4

# Depth Extraction from Unidirectional Integral Image data

Depth extraction is an important task both for content-based image coding and data manipulation, additionally a depth map is essential if real and/or computer generated objects are to be integrated within integral 3D TV images. However, due to the unique recording process involved in integral imaging, common depth estimation approaches cannot be directly applied to obtain depth from integral images. Previous work attempted to tackle the task used a novel depth extraction method based on image inverse. The method achieves good results on simulation but it is unable to obtain the depth solution from the real integral image due to the very ill-posed discrete correspondents associated with the direct process (Manolache1999, 2000, 2001). The theme of this PhD work is to find an alternative way to solve the depth extraction. For simplicity, unidirectional integral images are used. However, extension to Omni-directional integral images is straight forward.

# 4.1 Extracting viewpoint images from unidirectional integral image (UII) data

## 4.1.1 Characteristics of UII data

In chapter 2, it was shown that a true 3D optical model could be replayed from the recorded UII data. Although the data exists in a 2D format, it contains all the necessary 3D spatial information. The depth information is embedded in the recording in a unique manner. Prior to investigating the methods by which to extract depth information from UII data, a better understanding of the unique optical recording process is necessary.

Figure 4.1 is an example of UII data (Horseman). The 3-D scene can be replayed by overlaying it with a suitable micro-lens sheet.

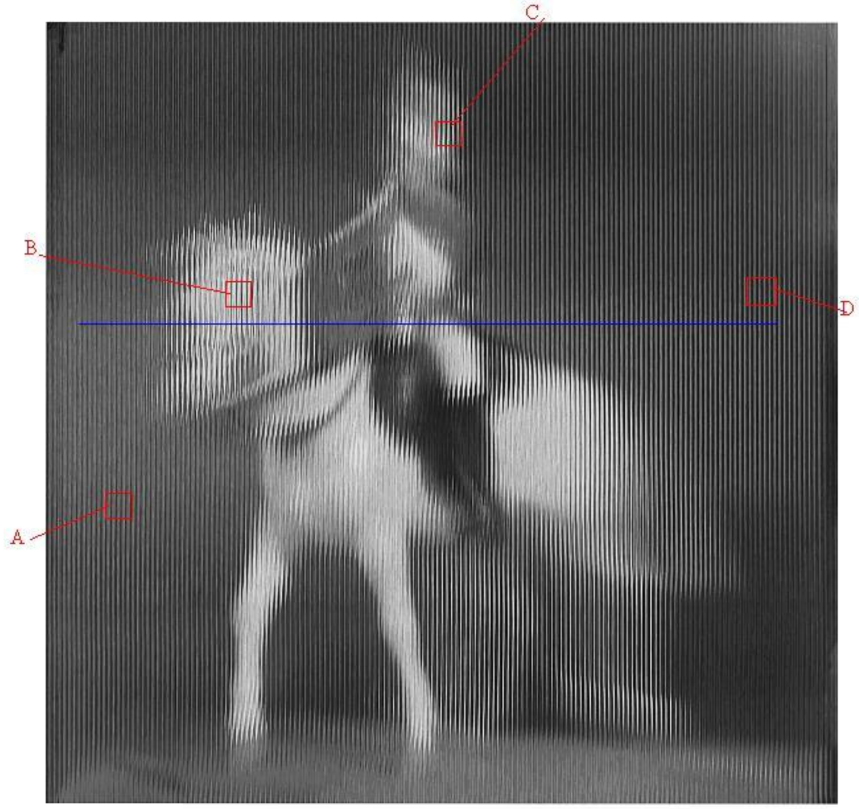


Figure 4.1: An example of captured UII data (Horseman).

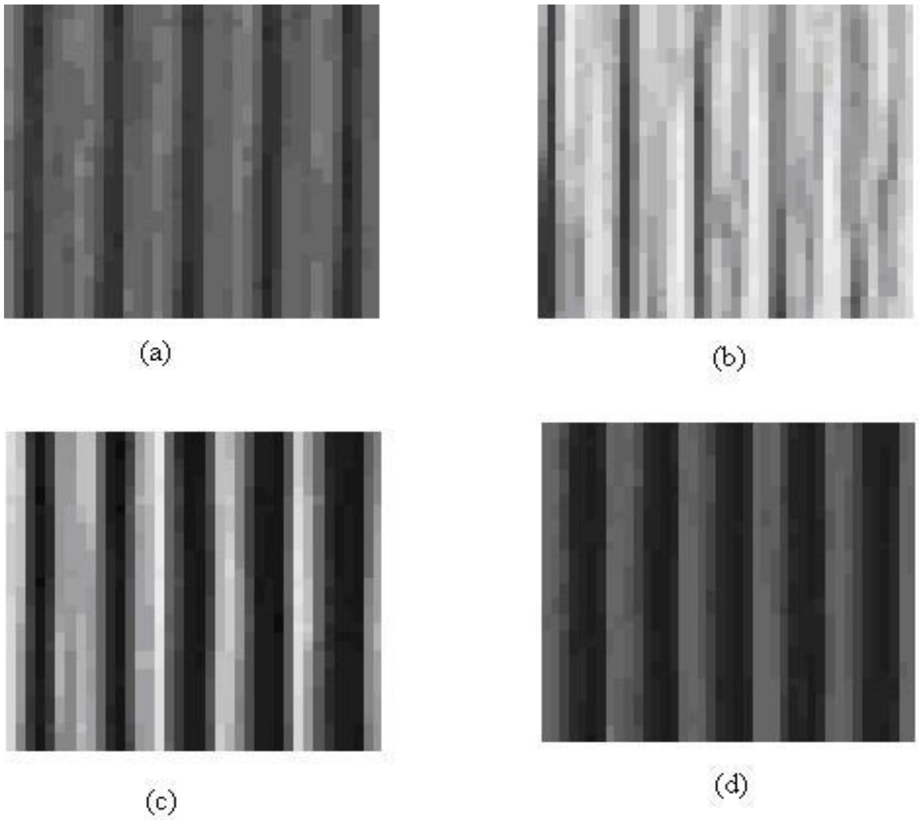


Figure 4.2: Four magnified sections of Figure 4.1. (a) region A, (b) region B, (c) region C, (d) region D.

Figure 4.2 (a)(b)(c)(d) illustrates four magnified areas taken from Figure 4.1. Observation shows that a regular banded structure exists in the image. Generally, the banded structure is comprised of two distinct components: black bands with low intensity and image bands with significant intensity variation. In this image, the width of the banded structure corresponds to 8 pixels.

Image profile analysis can be used to identify the intensity values along an outline path in an image. As an example, the profile along the blue horizontal line in Figure 4.1 is shown in Figure 4.3. A more detailed appreciation can be gained from Figure 4.4 that shows three enlarged portions of the data given in Figure 4.3. It can be seen from the profile that the low intensity values appear in the image at a regular interval. The interval is found to be at 8 pixels distance and is coincident with the banded structure observed from Figure 4.2.

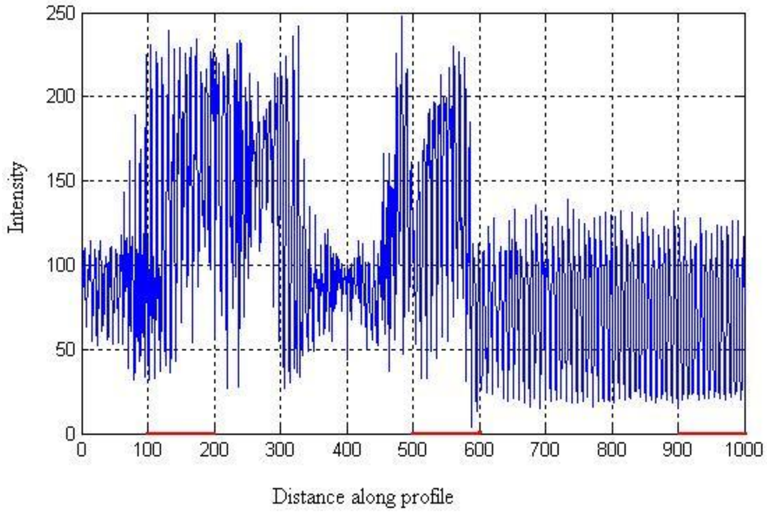
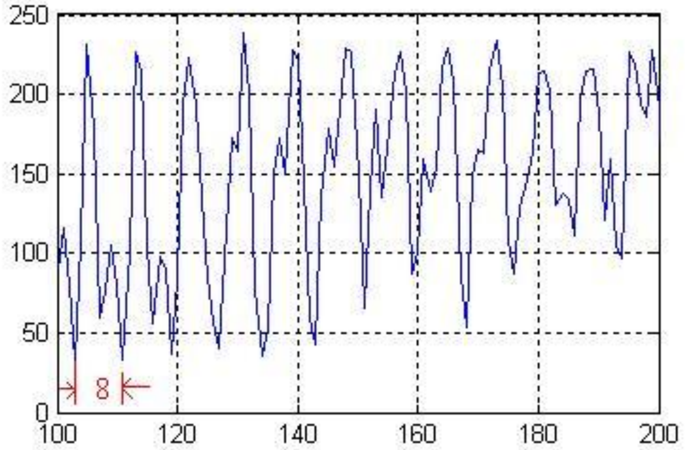
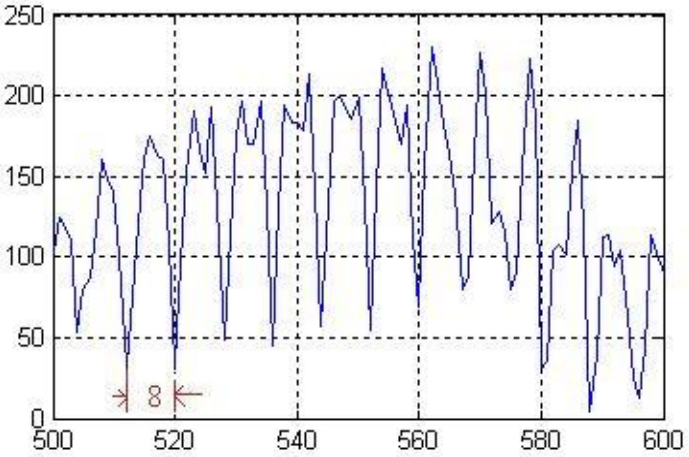


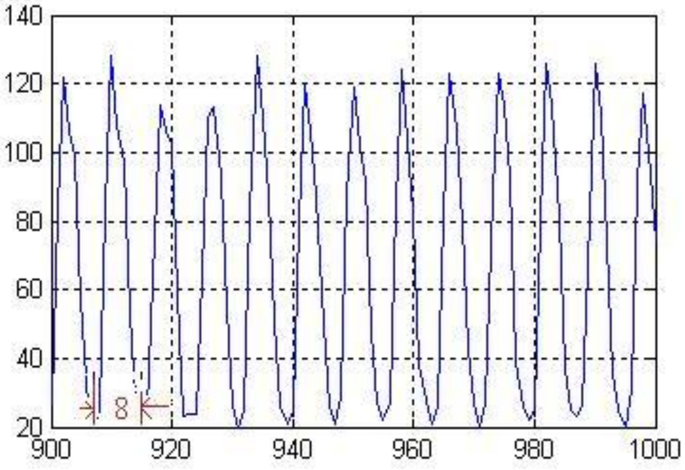
Figure 4.3: The profile of one horizontal line in Figure 4.1



(a)



(b)



(c)

Figure 4.4: Three enlarged parts of the image profile.

An auto-correlation analysis was further carried out along the horizontal line. The result is shown in Figure 4.5. It can be seen that the pixel at position P1 is more related to pixels at positions P9, P17 than pixels at the positions P2, P3. This is fundamentally different from the result obtained in an ordinary 2D image, where the neighbouring pixels have a high correlation due to being spatially adjacent. The local maximum of the correlation appears at a period of 8 pixels. This indicates that not only does the black band appear at the period of one microlens, the data within the image bands, which contains significant intensity variation has strong correlation at a particular interval. In this image, the interval is of one microlens pitch length.

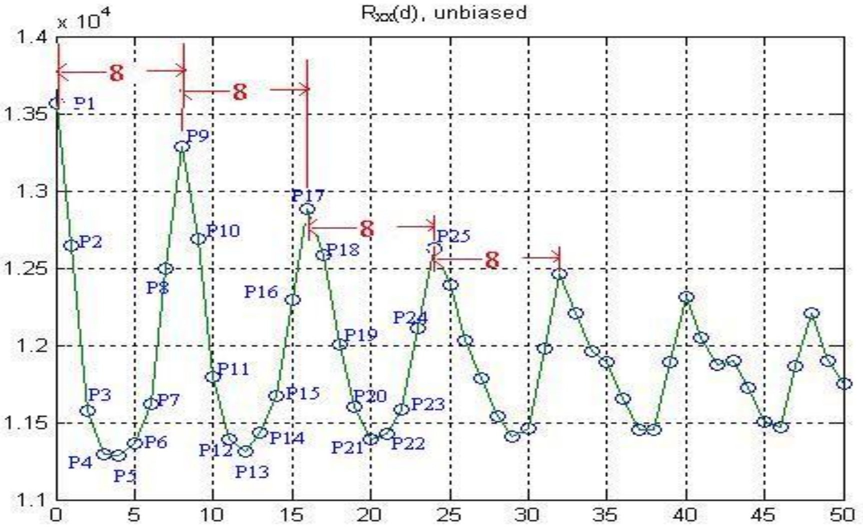


Figure 4.5: The unbiased auto-correlation analysis result. The intensity of a pixel in UII is closer to the pixels in one micro-lens distance rather than its adjacent pixels.

### 4.1.2 Extraction of viewpoint images from the UII data

The key feature of the II recording process is the use of a micro-lenses sheet to sample image data. For an ideal recording, all parallel rays entering the same micro-lens are recorded at the same position on the recording medium if it is coincident with the back focal plane of the micro-lens. This is shown in Figure 4.6(a). As a result, the rays in figure 4.6(a) annotated with an arrow are recorded at the local position marked as $n_1$, while the rays without arrows are recorded at the local position marked as $n_2$. Since many micro-lenses are involved in the integral image recording, for all the rays in the same direction, the recording pixels will have identical local positions under their own particular corresponding micro-lens. This is illustrated in Figure 4.6(b), where all rays annotated with an arrow are recorded on the positions marked by $n_1$. The different recording positions only depend on which micro-lens surface the ray reaches.

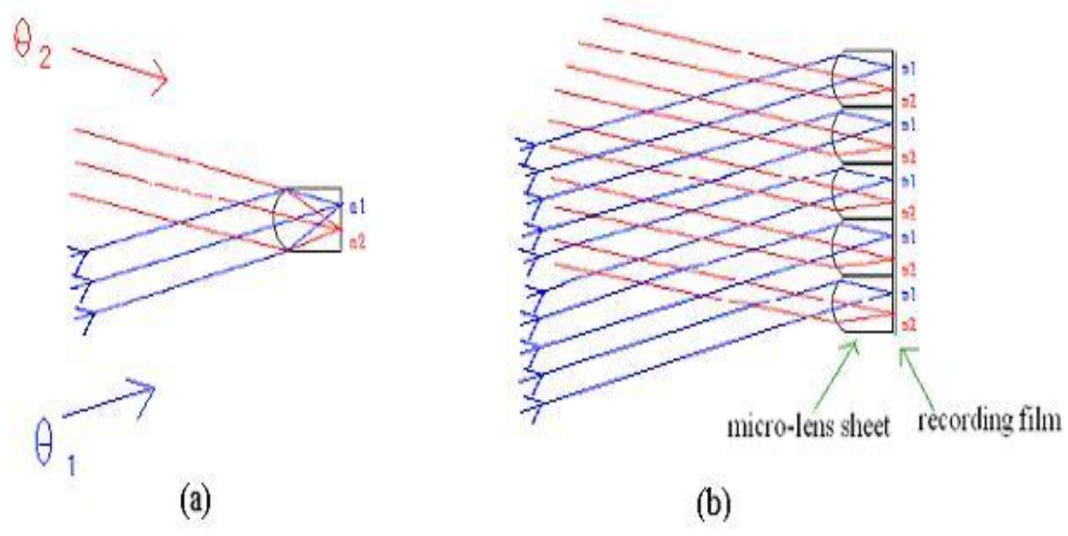


Figure 4.6: The direction selectivity in UII recording.

From the image point of view, all pixels in the UII data that are at the same relative position under different micro-lenses contain the recording of the object scene from a single direction. As an example, all pixels marked as $n_1$ contains only the recording from $\theta_1$ direction. Consequently, a strong correlation is found between pixels displaced by one micro-lens interval.

Therefore, by sampling all the pixels at the same local position under different micro-lenses, a new synthetic image can be formed. The new images, termed here viewpoint images, contain all the information within the object scene recorded from one particular view direction. By selecting pixels corresponding to other positions under the micro-lenses enables other viewpoint images to be constructed. Figure 4.7 graphically illustrates how the viewpoint images are extracted. The eight viewpoint images extracted from the captured UII data (Horseman) are shown in Figure 4.8.

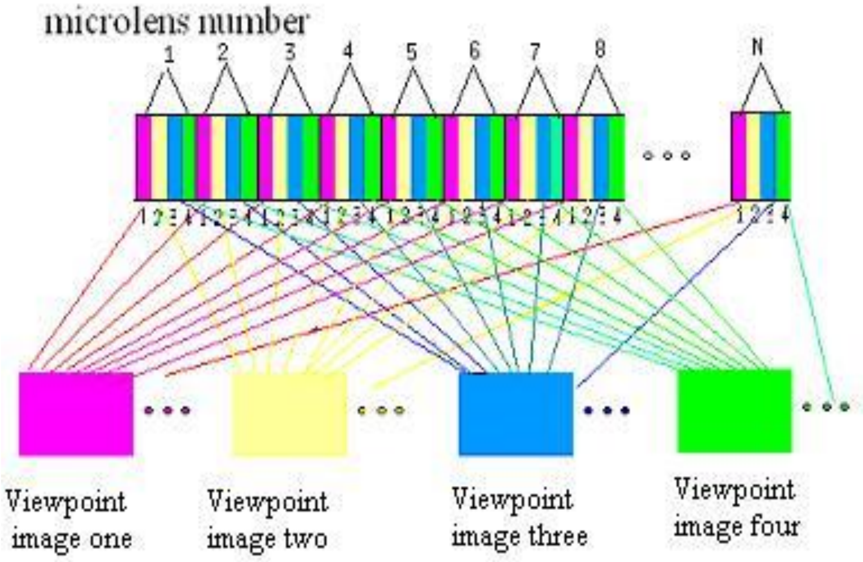
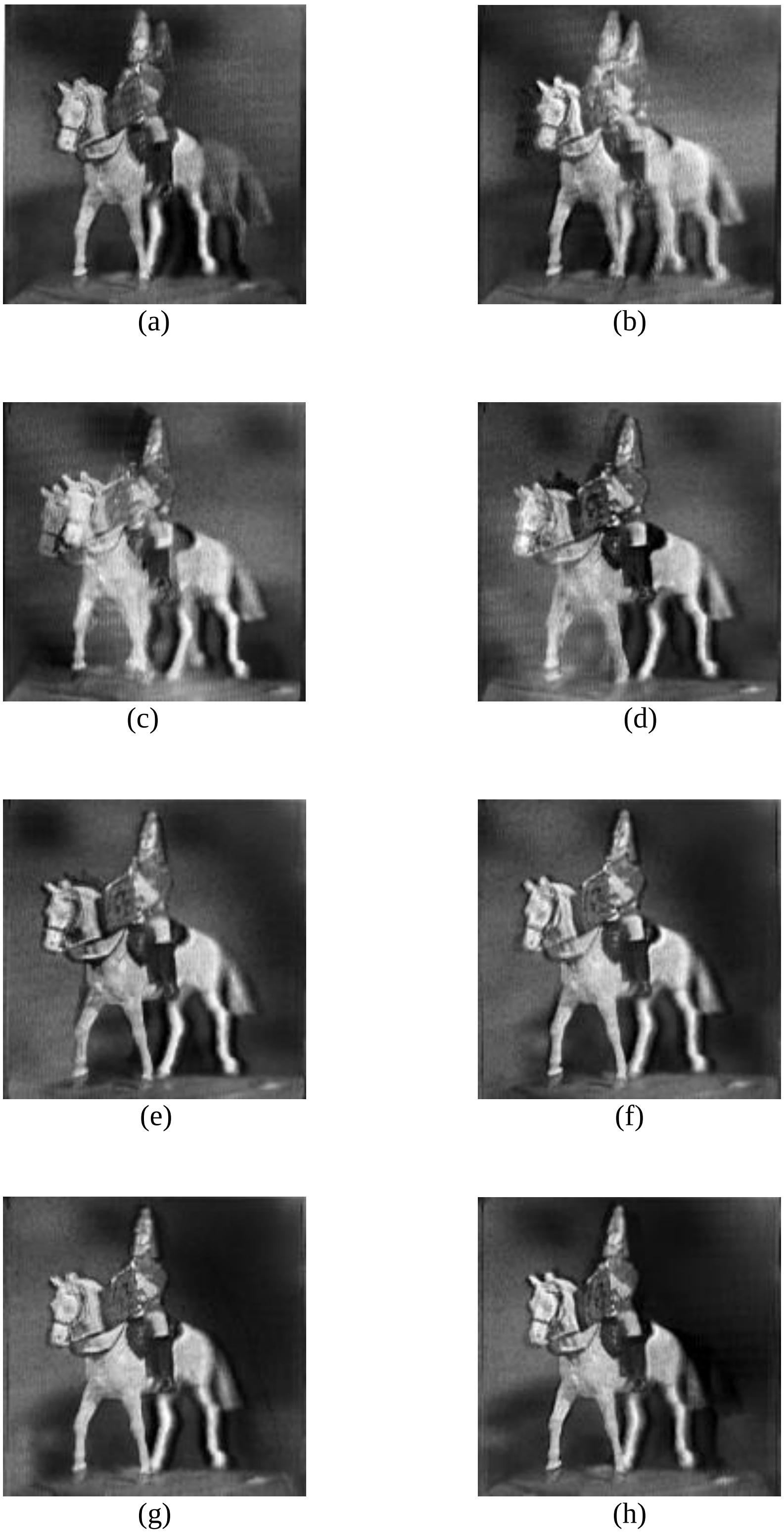


Figure 4.7: Illustration of viewpoint image extraction.

(For simplicity, assume there are only four pixels under each microlens, pixels in the same position under different micro-lenses, represented by the same color, are employed to form one viewpoint image.)

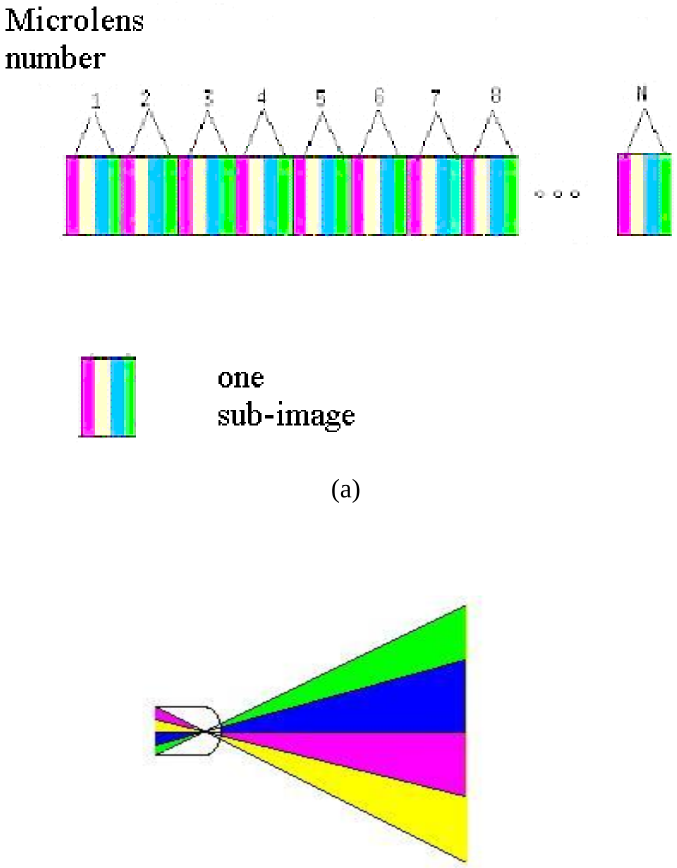
It is worth mentioning that the viewpoint image is different from the sub-image used in previous work, where a sub-image is defined as a group of adjacent pixels on the recording film under the same recording micro-lens, as shown in figure 4.9a. When look at each sub-image separately, each pixel of the sub-image is responsible for recording a large spatial angle, see figure 4.9b. This explains why high correlation does not exist between the spatially adjacent pixels.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.8: Eight Viewpoint images extracted from the UII data, Horseman.

(The image is scaled at 1/8 in the vertical direction)



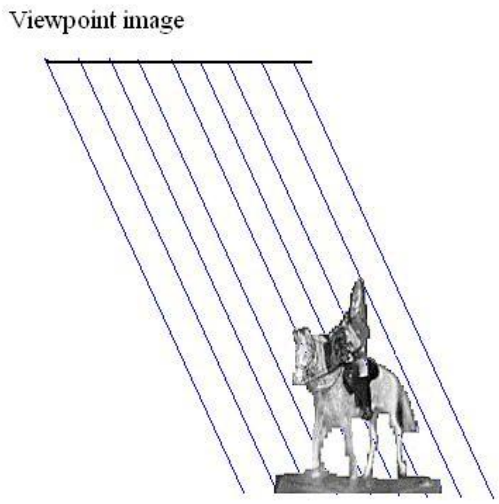
(a)

(b)

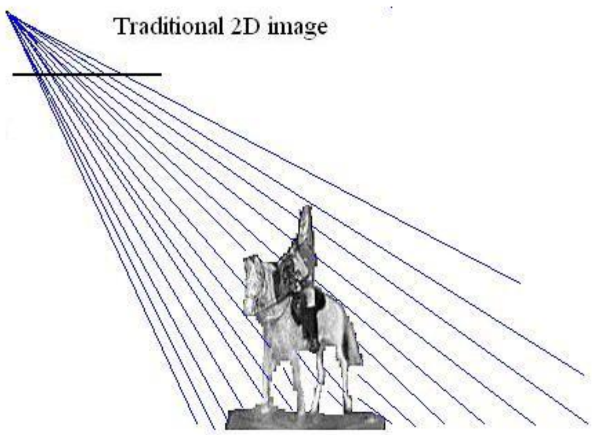
Figure 4.9: Illustration of the sub-image used in previous work.

(For simplicity, only four pixels are assumed under each micro-lens. A group of pixels under the same microlens are defined as one sub-image)

It is also worth noticing that the viewpoint image extracted is different from the traditional 2D image captured directly from an ordinary camera. It is a parallel projection recording of the 3D space rather than a perspective projection recording as in traditional 2D recording as shown in Figure 4.10. This illustrates how an integral display system is fundamentally different from a multi-view display system. In a multi-view display system, the multiple 2D images taken using traditional cameras placed at different positions are interlaced to construct a one multi-view image that generates a 3D effect replaying an angular disparity of the component images. However, the II is formed by using unique optical components (microlens sheet) encoding the true volume spatial optical model of the object scene directly in a form of a planar intensity distribution.



(a) The viewpoint image is a recording of the parallel projection of the 3D space



(b) The 2D image taken by a traditional camera is a recording of the perspective projection of the 3D space

Figure 4.10: The difference between a viewpoint image and a common 2D image

## 4.1.3 Pre-processing UII data before viewpoint image extraction

The UII data obtained through photographic capture is then scanned to the computer and converted to electronic data. The most common distortion caused by the scanner is that due to the rotational translation. This can be neglected in most of the image applications. However, in viewpoint image extraction, positioning is very important to enable the correct pixel to be extracted. Therefore, to correct for rotation and scale, the initial UII data obtained from the scanner is processed before extracting the viewpoint image (Forman 1999, Zaharia2001).

Two stages are involved in rotational correction. The first stage involves assessing the degree of rotational error in the scanned data. The second stage is to apply an appropriate correction in such a way that the structure of the intensity distribution is preserved.

Figure 4.11 diagrammatically illustrates a scanned UII data with an exaggerated rotational error. This can be recognized from the black bands (low intensity region) of the image, which should be vertical.

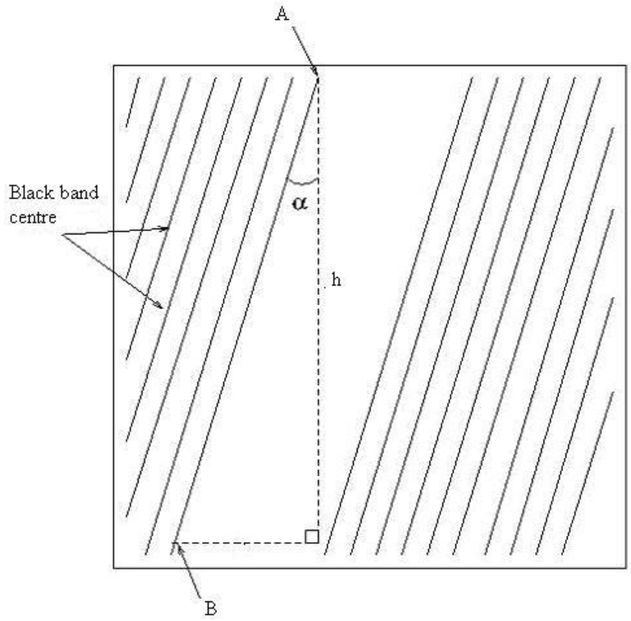


Figure 4.11: Rotational error in a scanned intensity distribution (Forman 1999)

If points ‘A’ and ‘B’ are chosen from a black band, the error angle can be calculated as:

$$\alpha = \tan^{-1}(\frac{w}{h}) \qquad 4.1)$$

To align the image, a rotation in the opposite sense ($-\alpha$ ) is necessary to perfectly align the UII data before further process.

Another important aspect is to make sure that all pixels are extracted from the same position under each micro-lenses. For simplicity, the integral images are resized so that an integral number of pixels (N) are presented under each micro-lens.

If p*(mm)* is the pitch size of the lenticular microlens sheet and *R(dpi)* is the scanning resolution, in order to have N pixels under each lenticular microlens sheet, the scale factor ($S_f$) can be calculated as:

$$S_f = \frac{N * 25.4}{pR} \qquad 4.2)$$

### 4.1.4 Application of viewpoint image extraction

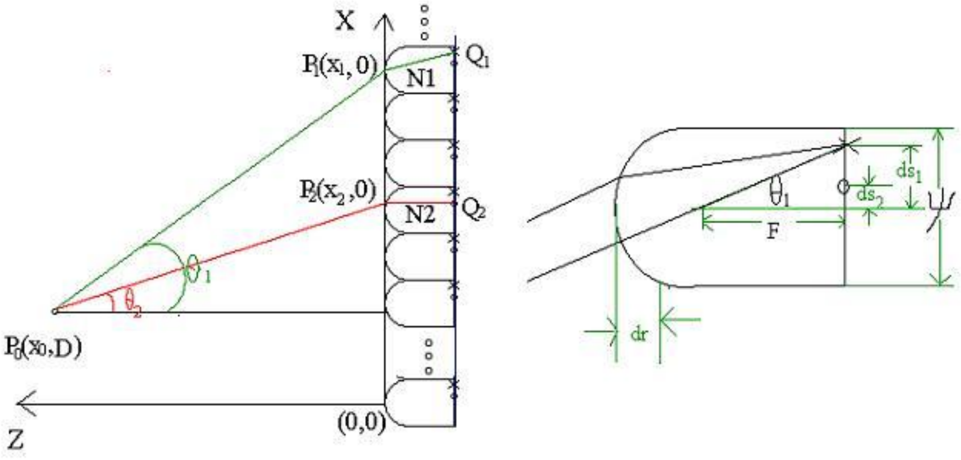
After viewpoint image extraction, the original integral image, which contains spatial coding information of the 3D object scene, is represented by a number of images in 2D recording format. This provides a possible way of analyzing and processing 3D integral images.

As an example, for image enhancement task, it is reasonable to expect the replayed 3D visual effect will be improved if the visual effect of each viewpoint image is improved. The idea is to first extract viewpoint images from the UII data and then apply existing 2D image enhancement algorithms to each extracted viewpoint image, thereby producing a high quality UII by integrating the enhanced viewpoint images. Similar processing can be carried out using other image processing and analysis tasks, such as, noise removal, edge detection. Consequently viewpoint images lead themselves to standard techniques in processing 3D UII data. In addition, the high correlation found within viewpoint images is also very useful in data compression in choosing an optimized scheme for pixel grouping. However, the main task in this thesis is to extract depth information from the UII data. Viewpoint image extraction provides a new approach to the task. The remainder of this thesis deals with this task in detail.

# 4.2 Geometric analysis of the UII recording process

1) Depth equation

To obtain the depth information from the extracted viewpoint images, geometric analysis of the optical recording procedure is carried out to find the mathematical relationship between object depth and the corresponding viewpoint image displacements. Figure 4.12 depicts the Cartesian coordinate system used in the analysis. Only one dimension is considered since only one direction disparity exists in the UII. The *z*-axis denotes the depth and the *x*-axis represents the lateral position. The *z*-axis starts from the plane coincident with the micro-lenses surface, while the *x*-axis is measured from the center of the first micro-lens.



(a) (b)

Figure 4.12: (a) The Cartesian coordinate used in geometric analysis. (b) An enlarged micro-lens. (Pixels marked by '*' are sampled out to form one viewpoint image, pixels marked by 'o' are sampled out to form another viewpoint image.)

Suppose the first viewpoint image is formed by choosing pixels at an offset $ds_1$ from micro-lens center, which record rays only from the $\theta_1$ direction; the second viewpoint image is formed by extracting pixels at an offset $\mathrm{ds}_2$ from the micro-lenses center, which records rays only from the $\theta_2$ direction. Consider an object point P $(x_0\ ,D)$ as shown in Figure 4.12. A ray from P in the $\theta_1$ direction forms the corresponding record in the first viewpoint image. This ray intersects the *x*-axis at point $\mathrm{P}_1(x_1\ 0)$ and then enters the micro-lens sheet at the $\mathrm{N}_1$th micro-lens. Following lens refraction, the ray is recorded at pixel *Q1.* *Q1* is the corresponding recording pixel in viewpoint image one. Similarly, the corresponding recording pixel $Q_2$ in the second viewpoint image is formed by project a ray from P in direction $\theta_2$. The ray enters the $N_2$th micro-lens, intersects the *x*-axis at point $P_2$ $(x_2\ 0)$ *a*nd is recorded on the film at pixel $Q_2$.

The following geometric relationship can be easily obtained from Figure 4.12:

$$x_1 = x_0 + D \cdot tg\theta_1 \qquad 4.3)$$

$$(N_1 - 0.5)\cdot\psi < x_1 + dr \cdot tg\theta_1 < (N_1 + 0.5)\cdot\psi \qquad 4.4)$$

$$tg\theta_1 = \frac{ds_1}{F} \qquad 4.5)$$

From equations 4.3)- 4.5), we have:

$$(N_1 - 0.5)\cdot\psi < x_0 + \frac{(D + d_r)\cdot ds_1}{F} < (N_1 - 0.5)\cdot\psi \qquad 4.6)$$

Similar equation can be written for the second viewpoint image:

$$(N_2 - 0.5)\cdot\psi < x_0 + \frac{(D + d_r)\cdot ds_2}{F} < (N_2 - 0.5)\cdot\psi \qquad 4.7)$$

Manipulating equation 4.6) and 4.7) yields:

$$(N_1 - N_2 - 1)\cdot\psi < \frac{(D + d_r)\cdot(ds_1 - ds_2)}{F} < (N_1 - N_2 + 1)\cdot\psi \qquad 4.8)$$

Defining 'baseline' $\Delta$=ds$_1$-ds$_2$ as the sampling distance between two viewpoint images. Since only one pixel is extracted from each micro-lens unit to form one viewpoint image, one pixel in one viewpoint image corresponds to one micro-lens unit in the UII. Therefore, one unit of disparity between the two viewpoint images is equal to one micro-lens unit distance in the original UII data, that is, $d = N_1 - N_2$.

Substituting *d* and $\Delta$ into equation 4.8), we have the mathematical relationship for depth D and disparity *d*:

$$(d-1)\cdot\psi < \frac{(D+d_r)\cdot\Delta}{F} < (d+1)\cdot\psi \qquad 4.9)$$

The depth equation of the object point $P_0$ depth (D) can be written as:

$$D = \frac{(d\pm 1)\cdot\psi\cdot F}{\Delta} - d_r \qquad 4.10)$$

Here, $d\pm 1$ represents that the expected value is *d* and varies in the range *d-1* to *d+1*. $\Delta$ is the sample distance between two viewpoint images. $\psi$, *F* is the pitch size, focal length of the micro-lens sheet respectively. $d_r$ is the height of the sag of the micro-lens, as shown in Figure 4.12 b).

In most cases, *dr<<D*, without concerning the error range, depth equation can be simplified to

$$D = \frac{d\cdot\psi\cdot F}{\Delta} \qquad 4.11)$$

Equation 4.11) can be used to calculate the object depth from the corresponding disparity between two viewpoint images. Given the corresponding displacement between two viewpoint images and the micro-lens sheet parameters used in recording, the depth of any object point can be easily calculated from the depth equation. It also can be seen that the displacement of viewpoint image pair is proportional to the depth and increases as the baseline increases.

2) Depth estimation error

The error range for the depth estimation from two viewpoint images is:

$$E = \frac{\psi}{\Delta}\cdot F \qquad 4.12)$$

The error range is in an inverse proportional to the baseline, which indicates that a longer baseline can give a better accuracy when using two viewpoint images in the depth estimation. Figure 4.13 graphically illustrate the ambiguity existing in the integral recording. Since all the parallel rays entering the same micro-lens have the same recording pixel, any point in the green diamond region in Figure 4.13a, can be identically recorded in the two viewpoint images as point P. Therefore, ambiguity exists when determining the exact position of the object point from two viewpoint images. Fortunately, the ambiguity region can be reduced by introducing more constraints, that is, by using the information from another viewpoint image, as shown in Figure 4.13b. In an ideal situation, this ambiguity region is reduced to zero by an infinite number of viewpoint images.

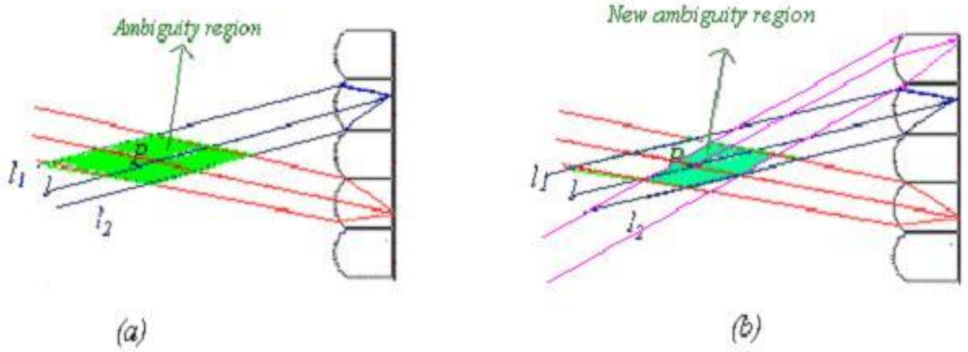


Figure 4.13: Ambiguity existing in integral recording: (a) The ambiguity region existing when two viewpoint images are considered; (b) The ambiguity region when more rays are involved.

## 4.3 Disparity analysis of viewpoint images

The derived depth equation shows that the depth can be calculated from the corresponding displacement between two viewpoint images. Therefore, the next step in depth extraction is to carry out a disparity analysis to establish the correspondence between two viewpoint images.

Disparity analysis has been a major research topic in computer vision for many years. For simplicity and effectiveness, a correlation-based block-matching method is used here. The basic idea of the block-matching method is to locate a candidate block in the second image that can best match a target block in the first image. This is illustrated in Figure 4.14, where only one dimension is considered for simplicity. Assuming two viewpoint images, $I_1$ and $I_2$, (x, y) are the coordinates of the point being analyzed. $I_1$(x, y) is the intensity of the point (x, y), *w* is the local window used in matching and R is the search range in the second image associated with the first image. In general, the algorithm can be mathematically described as finding out the matching position (x+d,y), $d \in [-R, R]$ in the second image where a score function has the minimum:

$$d^* = \arg\{\min_{d \in R}\{score(d)\}\} \qquad 4.13)$$

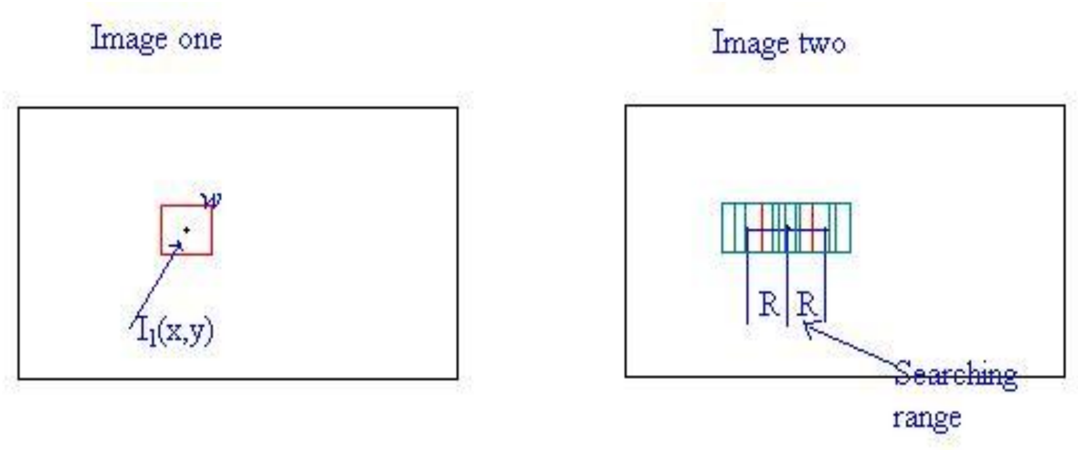


Figure 4.14: An illustration of block matching method.

Using three popular correlation-based block-matching criteria, namely, sum of the square difference (SSD), sum of absolute difference (SAD) and cross correlation (CC), the three score functions can be mathematically described as:

1. Criterion: SSD

$$score(d) = SSD(d) = \sum_{x,y \in w}[\hat{I}_1(x,y) - \hat{I}_2(x+d,y)]^2 \qquad 4.14)$$

2. Criterion: SAD

$$score(d) = SAD(d) = \sum_{x,y \in w}|\,\hat{I}_1(x,y) - \hat{I}_2(x+d,y)\,| \qquad 4.15)$$

3. Criterion: CC

$$score(d) = -CC(d) = \frac{-\sum_{x,y \in w}[\hat{I}_1(x,y) \cdot \hat{I}_2(x+d,y)]}{\sqrt{\sum_{x,y \in w}\hat{I}_1^2(x,y) \cdot \sum_{x,y \in w}\hat{I}_2^2(x+d,y)}} \qquad 4.16)$$

Where, $\hat{I}(x, y) = I(x, y) - \bar{I}$ and $\bar{I} = \frac{1}{N} \sum_{x,y \in w} I(x, y)$. N is the number of pixels within the window. The local window intensity adjustment is introduced in order to reduce the error caused by the variation of the illumination between different viewpoint images. This is found out to be particularly important in analysing the disparity between viewpoint images due to the directional illumination differences.

## 4.4 Experiments on depth estimation

To test the feasibility of depth estimation using the outlined approach, a UII, which contains a matchbox placed in front of a plane background as the scene was captured. Some letters are printed on the plane background as patterns to facilitate the disparity analysis. The recorded UII is therefore limited to two specific depth planes. The optical system used to capture the image is explained in detail in Davies and McCormick (Davies 1988). Figure 4.15 illustrates the integral recording process. The pitch size ($\psi$), focal length (F) and the radius of curvature (r) of the micro-lens array used in this recording are 0.6mm, 1.237mm and 0.88mm, respectively. The corresponding recorded UII data is shown in Figure 4.16. Figure 4.17 shows the 12 viewpoint images extracted from the UII data. The 3D scene can be replayed by overlaying the recording film with a micro-lenses sheet having the same parameters, as shown in Figure 4.18.

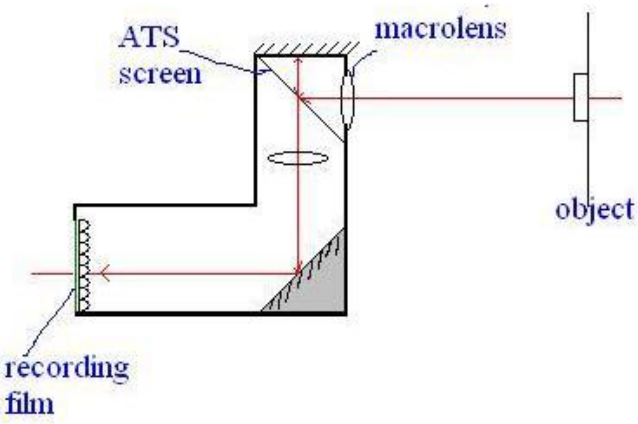


Figure 4.15: The Unidirectional Integral recoding process of a 3-D object, matchbox.

.

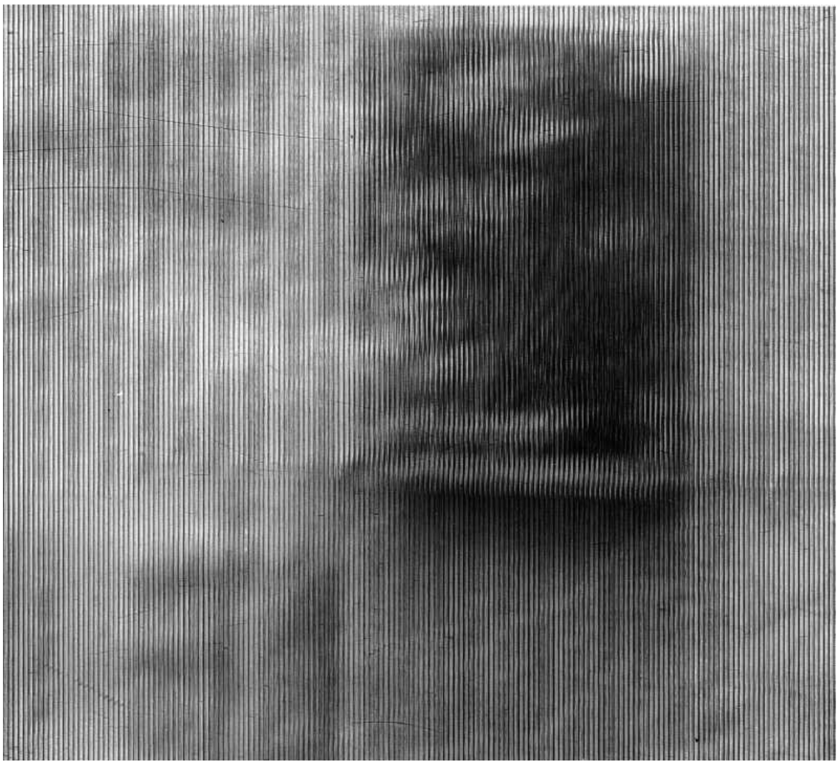


Figure 4.16: The captured UII data, matchbox.

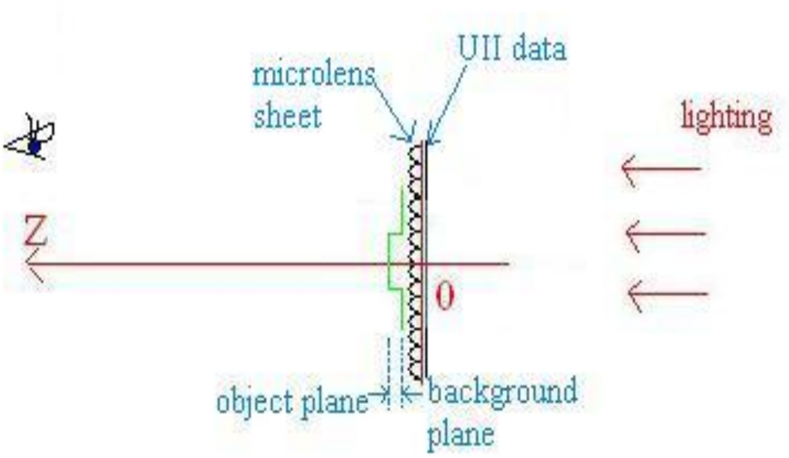
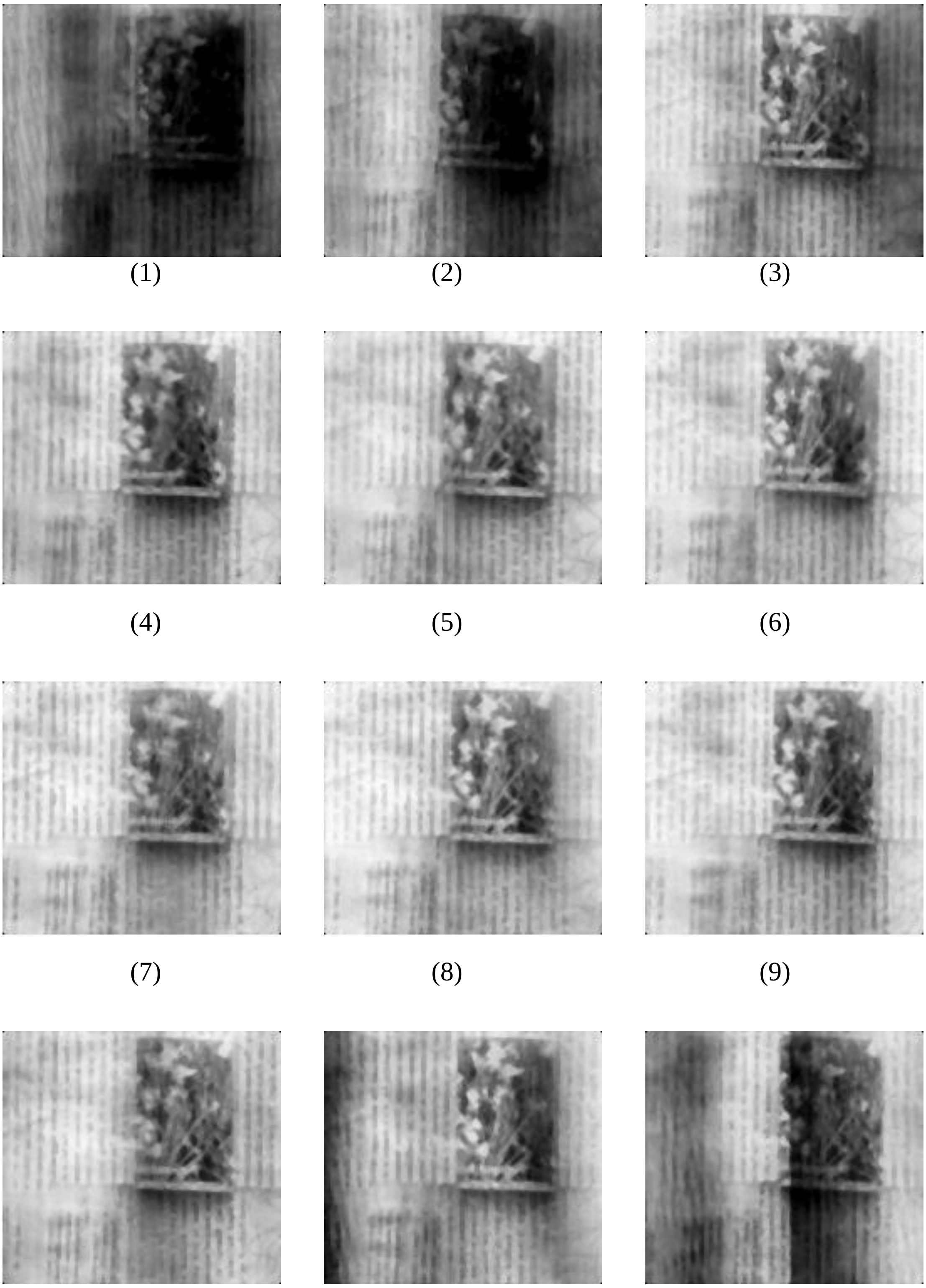


Figure 4.18: Replay of the 3-D optical object scene using a decoder (micro-lens array) with appropriate parameters. (For illustration, the replayed scene is shown in colour green).

.



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11) (12)

Figure 4.17: The twelve viewpoint images extracted from the UII data (matchbox).

Examining the twelve viewpoint images, it is seen that the first and last two are of poor visual quality. This is caused by recording quality deterioration near to the micro-lens edges (poor microlens form). As discussed in section 4.2, a long baseline is preferred in depth estimation to achieve high accuracy. Ignoring the poor viewpoint images, the 3rd viewpoint image ($VI_3$) and the 10th viewpoint image ($VI_{10}$) were selected to represent the longest practical baseline for the system and were initially used. Two local windows, one within object region (*w1*) and another within background region (*w2*), were chosen from the first image ($VI_3$) for measuring the depth, as shown in figure 4.19. The task is to find the corresponding matching positions in the second image ($VI_{10}$) for the two windows, respectively.

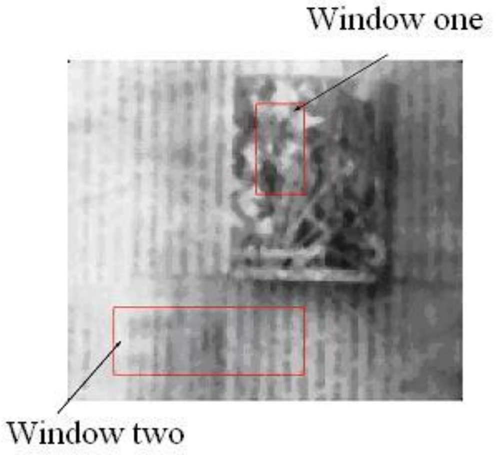
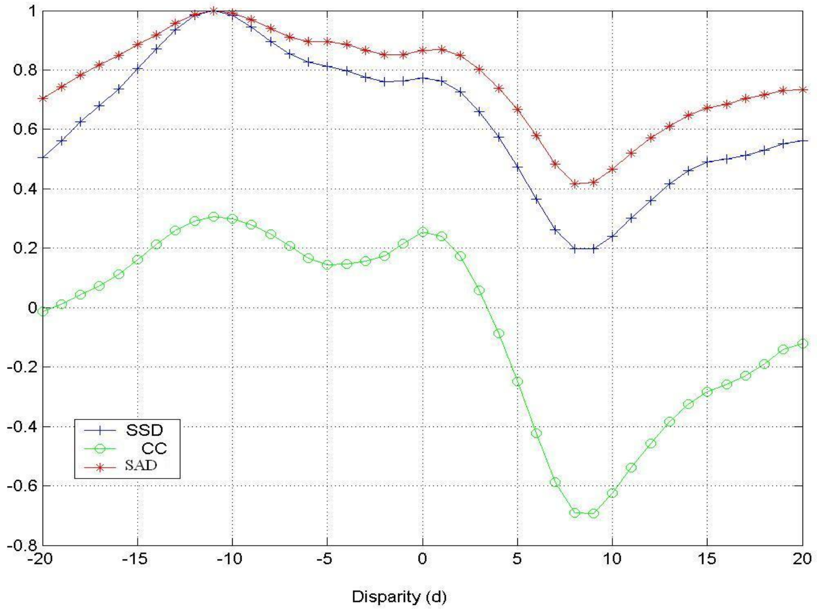
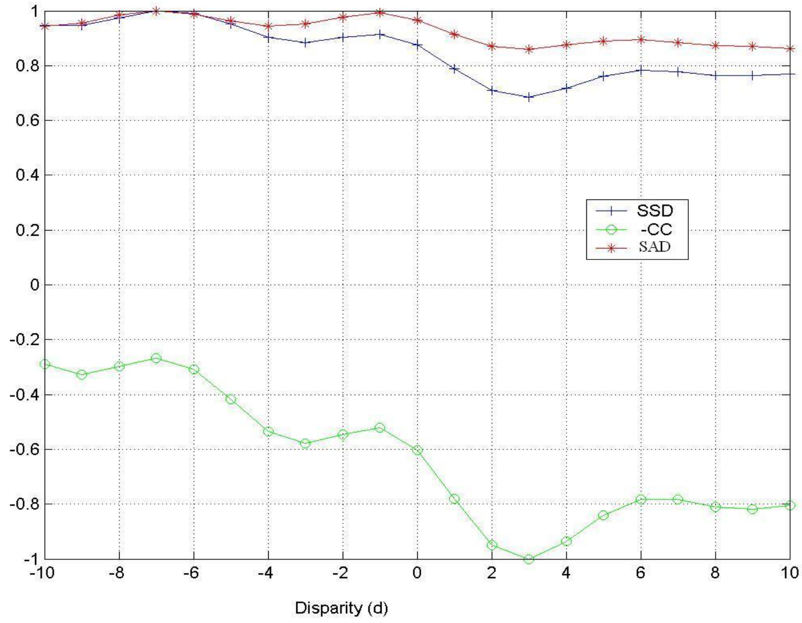


Figure 4.19: The two windows chosen for depth estimation.
[w*1*: (20, 80 - 60,100) *w2*: (110, 20 - 140, 100)]

A full-search algorithm was used sequentially adopting the three criteria previously described. The searching range is chosen from –20 to 20 pixels for the object region and –10 to 10 pixels for the background region. The results of the three score functions are plotted in Figure 4.20. For the chosen object region, a disparity of 9 is obtained from the SSD and CC criteria, 8 from the SAD criterion. Similarly, a disparity of 3 is obtained for the background region.



(a) *w1*



(b) *w2*

Figure 4.20: The score functions plotted for the two windows.

(All score functions are normalized to 1. The expected disparity is obtained from the position where the score functions have global minimum.)

Using the derived depth equation, the depth of the object region can be estimated as 19.1*mm* from the results of the SSD and CC criteria, 17.0*mm* on the result of the SAD criterion. From equation 4.12, the error range obtained from this viewpoint image pair (VIP) is 2.1*mm*. All three criteria give the depth of background region as 6.4*mm* with the same error range. Since the matchbox is attached to the plane background, the thickness of the matchbox is the difference between the two depths, as shown in figure 4.18. Therefore, the thickness of the matchbox is calculated as 12.7*mm*, 12.7*mm* and 10.6*mm* from the respectively three criteria. Since each measure gives an error range of *2.1mm*, the measure error range for the thickness of the matchbox is 4.2*mm*. Using a Vernier caliper gauge, the thickness of the matchbox is measured as 15.6*mm*. The relative error of the depth estimation obtained from the three criteria is 18.5%, 18.5%, 32%, respectively.

As a number of viewpoint images have been obtained from the UII data, it is reasonable to think that the result of the depth estimation can be improved by using more viewpoint images in the disparity analysis. Table 4.1 lists the disparities for the seven different viewpoint image pairs (VIPs) and Table 4.2 lists the corresponding depths calculated from each VIP. The corresponding error ranges for each VIP are list in Table 4.3. The statistical results of the depth obtained from seven different VIPs are given in Table 4.4

Table 4.1: Disparities obtained from different VIPs

| VIP | $VIP_1$<br>$VI_3$&$VI_4$ | $VIP_2$<br>$VI_3$&$VI_5$ | $VIP_3$<br>$VI_3$&$VI_6$ | $VIP_4$<br>$VI_3$&$VI_7$ | $VIP_5$<br>$VI_3$&$VI_8$ | $VIP_6$<br>$VI_3$&$VI_9$ | $VIP_7$<br>$VI_3$&$VI_{10}$ |
|---|---|---|---|---|---|---|---|
| SSD | 2 | 3 | 3 | 6 | 7 | 8 | 9 |
| CC | 2 | 3 | 3 | 6 | 7 | 8 | 9 |
| SAD | 2 | 3 | 3 | 6 | 7 | 8 | 8 |

(a) *w1*

| VIP | $VIP_1$<br>$VI_3$&$VI_4$ | $VIP_2$<br>$VI_3$&$VI_5$ | $VIP_3$<br>$VI_3$&$VI_6$ | $VIP_4$<br>$VI_3$&$VI_7$ | $VIP_5$<br>$VI_3$&$VI_8$ | $VIP_6$<br>$VI_3$&$VI_9$ | $VIP_7$<br>$VI_3$&$VI_{10}$ |
|---|---|---|---|---|---|---|---|
| SSD | 0 | 0 | 0 | 1 | 2 | 1 | 3 |
| CC | 0 | 0 | 0 | 1 | 2 | 1 | 3 |
| SAD | 0 | 0 | 0 | 1 | 2 | 2 | 3 |

(b) *w2*

Table 4.2: the depths (*mm*) estimated from different VIP's

| VIP | $VIP_1$<br>$VI_3\&VI_4$ | $VIP_2$<br>$VI_3\&VI_5$ | $VIP_3$<br>$VI_3\&VI_6$ | $VIP_4$<br>$VI_3\&VI_7$ | $VIP_5$<br>$VI_3\&VI_8$ | $VIP_6$<br>$VI_3\&VI_9$ | $VIP_7$<br>$VI_3\&VI_{10}$ |
|---|---|---|---|---|---|---|---|
| SSD | 29.7 | 22.3 | 14.8 | 22.3 | 20.8 | 17.8 | 19.1 |
| CC | 29.7 | 22.3 | 14.8 | 22.3 | 20.8 | 17.8 | 19.1 |
| SAD | 29.7 | 22.3 | 14.8 | 22.3 | 20.8 | 17.8 | 17.0 |

(a) *w1*

| VIP | $VIP_1$<br>$VI_3\&VI_4$ | $VIP_2$<br>$VI_3\&VI_5$ | $VIP_3$<br>$VI_3\&VI_6$ | $VIP_4$<br>$VI_3\&VI_7$ | $VIP_5$<br>$VI_3\&VI_8$ | $VIP_6$<br>$VI_3\&VI_9$ | $VIP_7$<br>$VI_3\&VI_{10}$ |
|---|---|---|---|---|---|---|---|
| SSD | 0 | 0 | 0 | 3.7 | 5.9 | 2.5 | 6.4 |
| CC | 0 | 0 | 0 | 3.7 | 5.9 | 2.5 | 6.4 |
| SAD | 0 | 0 | 0 | 3.7 | 5.9 | 4.9 | 6.4 |

(b) *w2*

Table 4.3: The error range (*mm*) obtained from different VIP's

| VIP | $VIP_1$<br>$VI_3\&VI_4$ | $VIP_2$<br>$VI_3\&VI_5$ | $VIP_3$<br>$VI_3\&VI_6$ | $VIP_4$<br>$VI_3\&VI_7$ | $VIP_5$<br>$VI_3\&VI_8$ | $VIP_6$<br>$VI_3\&VI_9$ | $VIP_7$<br>$VI_3\&VI_{10}$ |
|---|---|---|---|---|---|---|---|
| Error range | 14.8 | 7.4 | 4.8 | 3.7 | 3.0 | 2.5 | 2.1 |

Table 4.4: The statistical results of the depths (*mm*) estimated from VIP's

| Statistic results | Mean | Std | Median |
|---|---|---|---|
| SSD | 21.0 | 4.7 | 20.8 |
| CC | 21.0 | 4.7 | 20.8 |
| SAD | 20.7 | 4.9 | 20.8 |

(a) *w1*

| Statistic results | Mean | Std | Median |
|---|---|---|---|
| SSD | 2.6 | 2.8 | 2.5 |
| CC | 2.6 | 2.8 | 2.5 |
| SAD | 3.0 | 2.9 | 3.7 |

(b) *w2*

No obvious difference can be found from the statistical results of the three criteria. The results of the standard deviation show a slightly better estimate for the SSD and CC criteria. The mean value gives the thickness of the matchbox at 18.4*mm*, 18.4*mm* and 17.7*mm* for SSD, CC and SAD criterion, respectively. The median value from the three criteria gives the thickness at 18.3*mm*, 18.3*mm* and 17.4*mm*, respectively. Compared with the manually measured result (*15.6mm*), all results have an estimation error less than 20%

## 4.5 Summary

In this chapter, a method of extracting depth information from the UII data by extracting viewpoint images is explored and discussed. Three steps are involved in the approach: (i) Extracting viewpoint images from UII data; (ii) Finding the displacement from the extracted viewpoint images; (iii) Calculating the depth from the displacements.

The viewpoint image is formed by sampling pixels at the same local position under different micro-lenses. Each viewpoint image is a 2D parallel projection of the 3D scene. Through geometrically analyzing the UII recording process, a depth equation has been derived which describes the mathematical relationship between object depth and the corresponding viewpoint images displacement. Using the depth equation developed, the task of depth estimation has then converted into the task of disparity analysis. A correlation-based block matching method has been chosen to find the disparity among viewpoint images.

To test the efficiency of the approach, an object scene which only contains two depths (one for the object surface, one for the background) has been captured as an integral image. By selecting two matching windows in the object region and background region, respectively, the thickness of the object has been estimated with an error of less than 20%.

The performance of the disparity analysis is of great importance in achieving correct depth estimation. The following work is mainly concerned with developing algorithms to improve the performance of disparity analysis on extracted viewpoint images.

# Chapter5

# Disparity Analysis on Extracted Viewpoint Images

After obtaining the depth equation in Chapter 4, the task of depth extraction from UII data has successfully reduced to the task of disparity analysis on extracted viewpoint images. The performance of the depth estimation directly depends on the performance of the disparity analysis on viewpoint images. It was therefore decided to investigate approaches in order to improve the disparity analysis. This can be divided into two sections. In the first section, a multi-baseline stereo technique is adapted to improve the matching results by using the information from multiple viewpoint images. Subsequently, an algorithm based on a neighborhood constraint & relaxation technique to further improve the disparity analysis performance is presented. Experiments show that both techniques work effectively in achieving a good matching result. An obvious improvement in both precision and correctness can be found from the depth measurement results. The depth map and object space reconstruction can be achieved from the UII data with acceptable quality, both on the captured and computer generated UII data.

# 5.1 Disparity analysis using a multi-baseline technique

## 5.1.1 Trade-off between precision and correctness in matching

The derived depth equation indicates that the displacement for an object point between recording viewpoint images is proportional to the baseline of the viewpoint image pair. The baseline acts as a magnification factor when measuring disparity (*d)* from which the depth (*D)* can be obtained. The depth equation also reveals that the accuracy of the depth estimation is related to the baseline. For precise distance estimation, a long baseline is desired. However, a longer baseline means a larger disparity range must be searched to find the matching position. As a result, the chance of a false match is greater when using a long baseline. This is very similar to stereo matching in which choosing a baseline is a tradeoff between precision and accuracy.

To examine the efficiency of the matching strategy in choosing a baseline, a simple experiment was undertaken. Using the example on the UII data (matchbox) in the pervious chapter where a window *w2* within the background region is chosen for measuring the background depth in the scene. Using the viewpoint image pair ($VIP_7$) with the longest baseline, a disparity of 3 is obtained within the disparity search length $R_7$ ($R_7$=10), see Figure 4.19 (b). This gives the depth of the background as $6.4mm$ with an error range of 2.1*mm*. Now, considering a relatively small sized window (*w3)* within the background region, as shown in Figure 5.1, the depth is again calculated. Figure 5.2 shows the corresponding evaluation function for the window (*w3*) from the same viewpoint image pair. This time, a matching position of 9 can be found on the result, which gives the depth at 19.2mm for the region. This is quite different from the result obtained using *w2*. However, the result is expected to be the same as *w2* since *w3* and *w2* refer to regions at the same depth.

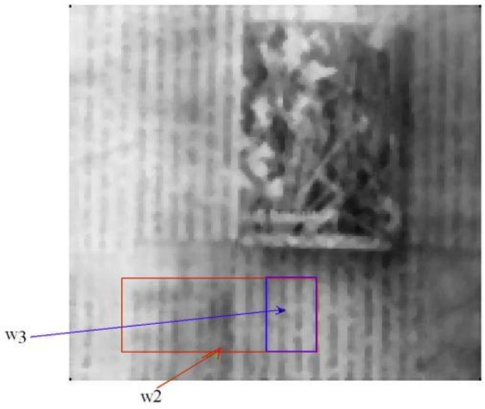


Figure 5.1: Use of a relatively small sized window (*w3*)

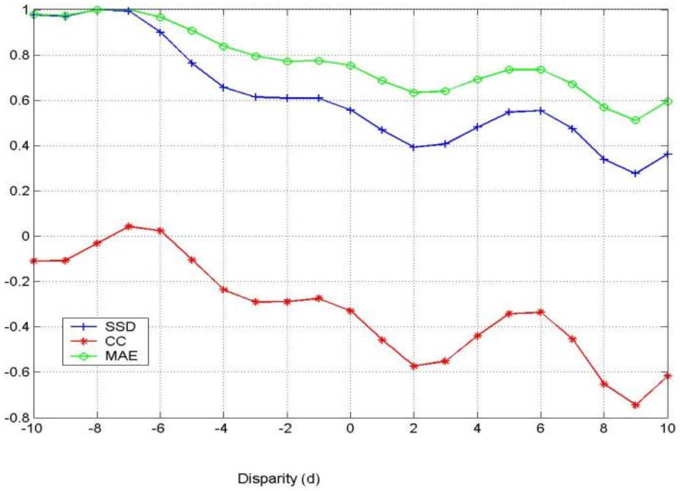
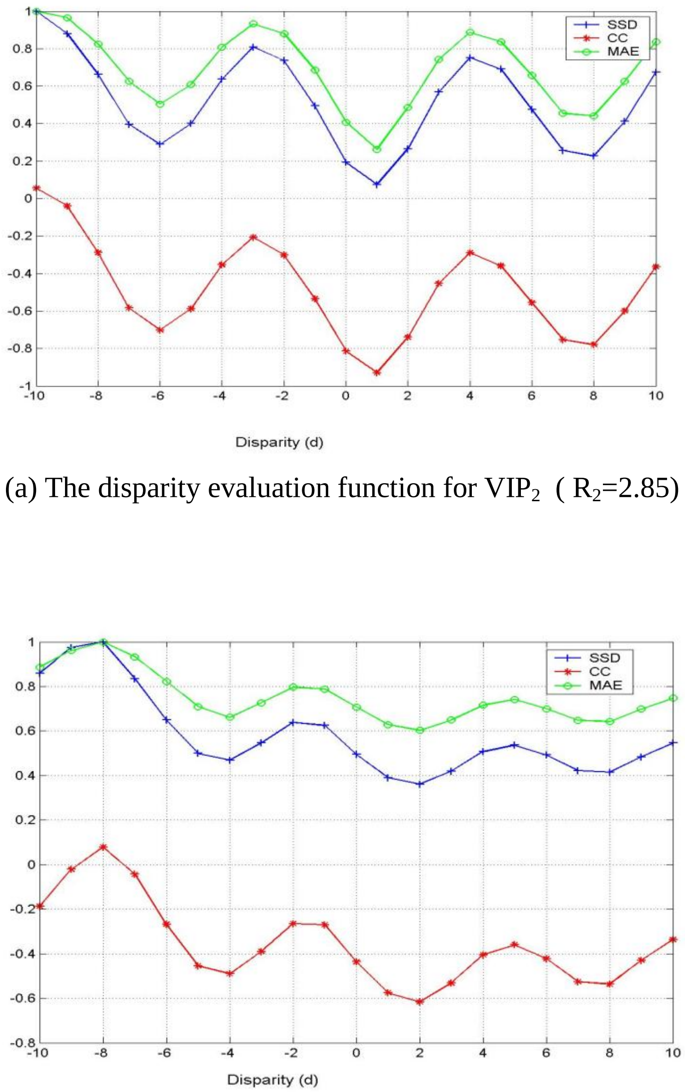


Figure 5.2 The disparity evaluation function for *w3* on the VIP$_7$ ( R$_7$=10)

This different depth estimation result is caused by false matching in the disparity analysis. False matching occurs when the matching signal is not strong enough to overcome the interference from the noise signal. A relatively small sized window is more easily affected by noise since less matching information is available. If the disparity evaluation functions obtained on w2 and w3 (Figure 4.20 and Figure 5.2) are considered it is seen that several local minima exist on the curve. The appearance of the local minimum in the evaluation function is caused by the periodic intensity distribution in the scene which can be observed from figure 5.1. False matching is especially easy to occur around this periodic position.

The false matching occurs could be precluded by using a short baseline since a relatively shorter disparity searching range would be presented. For illustration, Figure 5.3 show the disparity evaluation functions obtained from two other viewpoint image pairs with shorter baseline: VIP$_2$ and VIP$_4$. The baseline in VIP$_2$ is 2/7 of that in VIP$_7$. For the same depth search range, the corresponding disparity searching range is 2/7 of that in VIP$_7$ since the disparity is in a direct ratio to the baseline of the image pair. Previously, the disparity search range for VIP$_7$ (R$_7$) is chosen as 10, therefore, $R_2 = \frac{2}{7} * R_7$ =2.85. Similar relationship exist for VIP$_4$ which has a baseline length of 4/7 of the VIP$_7$, and therefore has the corresponding disparity $R_4 = \frac{4}{7} * R_7$ =5.71. Within the corresponding searching range, only one local minimum can be found on the corresponding functions, it is 1 for VIP$_2$ and 2 for VIP$_4$ respectively, see Figure 5.3. The corresponding depth can be calculated as 7.4*mm* in both cases. The error range obtained for the two VIP is 7.4*mm* and 3.7*mm* respectively, see Table4.3. These results ($7.4 \pm 7.4mm$ and $7.4 \pm 3.7mm$) are consistent with the result previously obtained on window *w2* ($6.4 \pm 2.1mm$) within the error range. However, the ambiguity we had from the two viewpoint image pairs are 7/2 and 7/4 times of using the longest baseline and therefore the uncertainty in depth estimation is increased.

The trade-off in choosing baseline is similar to what happens in stereo matching, where two common approaches are used to deal with the problem. The first approach is to use a coarse-to-fine strategy where a low resolution image is used at the first stage to obtain a coarse matching result. A fine matching is then carried out on the high resolution image using the results obtained from the coarse matching (Bierling 1988, Barnard 1989, Hannah 1989, Chen and Medioni 1990). An alternative approach is to use multiple consecutive images and take advantage of the redundancy of information contained in multiple images of the same scene to obtain a robust and precise measurement. The tracking of features becomes easy between a pair of consecutive images over a short searching range. A more precise estimation can be obtained by integrating the noisy individual measurements imposing certain constraints (Matthies 1989, Okutomi 1993). The procedure using a coarse resolution does not always result in the removal of the false matching, especially when it is caused by the inherent ambiguity Therefore, an approach of using multiple images of the same scene is considered since multiple viewpoint images with regular baseline interval have been obtained from the UII data by viewpoint image extraction. More precisely, a multi-baseline technique developed by Okutomi and Kanade is adapted (Okutomi 1993). The multiple-baseline stereo algorithm developed by Okutomi and Kanade and the modification made to adapt to viewpoint images extracted from integral images are described in the following parts of this section.



(a) The disparity evaluation function for VIP$_2$ ( R$_2$=2.85)

(b) The disparity evaluation function for VIP$_4$ ( R$_4$=5.91)

Figure 5.3: Using a shorter baseline in matching analysis on window *w3*

## 5.1.2 Using a multi-baseline technique in depth analysis

1) Multi-baseline stereo algorithm

In the multiple-baseline stereo algorithm developed by Okutomi and Kandade (Okutomi 1993), the multiple stereo image pairs with different baselines are generated by a lateral displacement of a camera. The matching is performed by accumulating the evaluation function from each stereo image pair and then making a judgment from the accumulated evaluation function. This is different from the traditional method, where the judgment is carried out on the score function from each pair directly. The final result is obtained from all intermediate judgments.

For illustration, if a one dimensional situation is considered and SSD criterion is used as the score function, the traditional stereo matching algorithm can be described as: Given a stereo pair, find the disparity that minimizes the score function.

$$d^* = \arg\{\min_{d \in R}\{score(d)\}\}$$

$$score(d) = SSD(d) = \sum_{x,y \in w} [\hat{I}_1(x,y) - \hat{I}_2(x+d,y)]^2 \qquad 5.1)$$

The Multi-baseline algorithm in stereo is described as: Given n stereo pairs with different baselines, find the ζ that minimizes the sum of the SSD-in-inverse-distance function.

$$\zeta = \frac{1}{D} \qquad \zeta^* = \frac{1}{D^*}$$

$$\zeta^* = \arg\{\min_{\zeta \in R'}\{score(\zeta)\}\}$$

$$score(\zeta) = SSSD(\zeta) = \sum_{i=2}^{n} \sum_{x,y \in w} [\hat{I}_1(x,y) - \hat{I}_i(x + \Delta_i F\zeta, y)]^2 \qquad 5.2)$$

In the new matching score function, the inverse distance $\zeta$ ($\zeta = \frac{1}{D}$) is used as the variable instead of the disparity *d*. The reason for using $\zeta$ instead of *d* is because the disparity *d* varies as the baseline varies in stereo vision, while the depths kept constant for different baselines. Therefore, the inverse distance does not change with different baselines and the same minimum can be expected to obtain from the SSD-in-inverse distance functions with respect to different baselines. This enables the score function from different image pairs to be directly accumulated.

The "multiple-baseline stereo" algorithm has proved to be effective in reducing mismatching and increase matching precision in stereo matching by both mathematical analysis and practical work in stereo vision (Okutomi 1993).

2) A modified multi-baseline algorithm

In the current work, disparity analysis from extracted viewpoint images is the objective. The extracted viewpoint images form multiple VIPs with different 'baseline' at regular length intervals. In this respect, the multiple VIPs can be viewed as being similar to multiple stereo image pairs used in multi-baseline stereo. However, the viewpoint images are generated using a different approach and the definition of baseline is different from that used in stereo vision as well. Therefore, some modification to the standard algorithm is necessary. The modification proposed is to use a SSSD-in-distance function (*SSSD(D))* to replace the SSSD-in-inverse-distance function ($SSSD(\zeta)$) used in stereo vision.

The modified multi-baseline algorithm can be written as:

$$D^{*} = \arg\{\min_{D\in R'}\{score(D)\}\}$$

$$SSSD(D) = \sum_{i=2}^{n}\sum_{x,y\in w}[\hat{I}_1(x,y) - \hat{I}_i(x + \frac{D\cdot\Delta_i}{\psi\cdot F}, y)]^2 \qquad 5.3)$$

The use of $D$ instead of $\zeta$ is because the depth is directly proportional to the disparity in the integral depth equation rather than an inverse relationship in stereo vision. This allows the evaluation functions from different viewpoint image pairs can be easily accumulated.

3) Mathematical proof of the modified multi-baseline algorithm

The mathematical proof on how the ambiguity can be removed from multiple viewpoint image pairs using the modified multi-baseline criterion is now outlined:

I) Suppose $I_1(x,y)$ and $I_i(x,y)$ are the intensity functions of an image pair. The image intensity functions $I_1(x,y)$ and $I_i(x,y)$ near the matching position can be represented as:

$$I_1(x,y)=I(x,y)+n_1(x,y)$$
5.4)

And

$$I_i(x,y)=I(x,y-dr_i)+n_i(x,y) \qquad 5.5)$$

Where $n_1(x,y)$ and $n_i(x,y)$ are used to represent the noise and $dr_i$ is the disparity for the image pair on matching point.

In a traditional matching algorithm, the SSD value of pair $i$, $SSD_i\ (d_i)$ over a window $w$ at a pixel position $(x,y)$ for the candidate disparity $d_i$ is defined as

$$SSD_i(d_i) = \sum_{x,y\in w}[I_1(x,y) - I_i(x+d_i,y)]^2 \qquad 5.6)$$

The $d_i$ that gives a minimum of $SSD_i(d_i)$ is determined as the estimate disparity.

Assuming $n_i(x,y)$ is independent Gaussian white noise $n(x,y) \sim N(0,2\sigma_n^2)$, the expected value of $SSD_i(d_i)$ can be represented as

$$E[SSD_i(d_i)] = \sum_{x,y\in w}[I\ \ (x,y) - I(x+d_i-dr_i,y)]^2 + 2N_w\sigma_n{}^2 \qquad 5.7)$$

Where, $N_w$ is the number of points in the searching window.

$E[SSD_i(d_i)]$ is expected to take a minimum value when $d_i$ is at the right disparity $dr_i$.

II) Suppose there is a same or similar pattern on the image around the matching position *(x,y)* so that

$$I(x,y)=I(x+a,y) \quad 5.8)$$

where, $a\neq 0$ is a constant. This will cause ambiguity since $E[SSD_i(d_i)]$ is also expected to be a minimum at position $dr_i+a$.

III) In our current task, from the derived depth equation in chapter 4, we know that the measured disparity *d* depends upon the depth of object point (*D*) and the baseline of the viewpoint image pair. Therefore, the candidate and real disparity $d_i$, $dr_i$ can be given as:

$$d_i = \frac{D\cdot\Delta_i}{F\cdot\psi} \quad 5.9)$$

and

$$dr_i = \frac{D_r\cdot\Delta_i}{F\cdot\psi} \quad 5.10)$$

where *Dr* and *D* are the real and candidate depth of point *(x,y)*, respectively.

Substituting (5.9) and (5.10) in (5.7), the $E[SSD_i(D)]$ with respect to the distance D is obtained as.

$$E[SSD_i(D_i)] = \sum_{x,y\in w}[I(x,y)-i(x+\frac{(D-D_r)\cdot\Delta_i}{F\cdot\psi},y)]^2 + 2N_w\sigma_n^2 \quad 5.11)$$

IV) Now again suppose *I(x,y)* has the same pattern around *(x,y)* and *(x+a, y)*. The equation 5.11 is expected to have a minimum value both at $\frac{(D-D_r)\cdot\Delta_i}{F\cdot\psi}=0$ and $\frac{(D-D_r)\cdot\Delta_i}{F\cdot\psi}=a$, which corresponding to *D=Dr* and $D=D_r+\frac{a\cdot\psi_i\cdot F}{\Delta_i}$. It is important to notice that the false depth estimation, $Df_i=\frac{a\cdot\psi\cdot F}{\Delta_i}$, varies for different baseline.

In the modified multi-baseline algorithm, the new evaluation function, *SSSD(D),* is now the sum of *SSD(D)*:

$$SSSD(D)=\sum_{i=1}^{n}[SSD_i(D)]$$

The expect value of *SSSD(D)* is:

$$E[SSSD(D)]=\sum_{i=1}^{n}\sum_{x,y\in w}[I(x,y)-I(x+\frac{(D-D_r)\cdot\Delta_i}{F\cdot\psi},y)]^2+2nN_w\sigma_n^2 \qquad 5.12)$$

This function is expected to have minimum only when all *SSDᵢ(D)* have the minimum.

V) Now, considering the ambiguity caused by the same pattern around *(x, y)* and *(x, y+a)*. Each *SSDᵢ(D)* can achieve the minimum at two positions( *D=Dr*, *D=Dr+* $\frac{a*F*\psi}{\Delta_i}$). However, the false depth estimation *D=Dr+* $\frac{a*F*\psi}{\Delta_i}$ varies for different baseline since $\Delta_i$ varies with the change of *i*. The equation achieves a minimum for every *SSDᵢ(D)* only when *D=Dr*. Therefore, even when ambiguity exists in the object space, *Dr* is the only position that $E[SSD_i(D)]$ can achieve minimum. Ambiguity is removed using the modified multi-baseline algorithm.

## 5.1.3 Experiments

### 5.1.3.1 Reducing false matching

An experiment was carried out to show how the false matching in depth estimation can be removed using the modified multi-baseline algorithm on multiple viewpoint images:

Figure 5.4(h) shows the accumulated evaluation function obtained from the modified multi-baseline algorithm on the window *w3* used in section 5.1.1, see Figure 5.1.  For comparison, the disparity evaluation functions from each viewpoint pair with respect to different baselines are shown in Figure 5.4 (a)-(g).

The results of figure 5.4 can be summarized as:

(1) A longer baseline produces better resolution but the chance of mismatching is greater since a longer searching range is involved. As can be seen from Figure 5.3(d)-(f),  more than one local minimum exist in the valid searching range and mismatching can occur around those positions due to the effect of noise or a non-balanced illumination, as an example, the false matching can be obtained on Figure 5.4(g).

(2) Mismatching can be precluded from the image pairs using a shorter baseline where only one local minimum exist in the score function, see Figure 5.4(a)-(c). However, the curve around the minimum is rather flat which gives a low resolution in depth estimation.

(3) By considering all the image pairs with different baselines, the accumulated score function gives a single clear, sharp minimum, see Figure 5.4(h).  The mismatching has been effectively removed by integrating the results using different baselines.

The outcome is matching with the mathematical analysis described in previous sections.

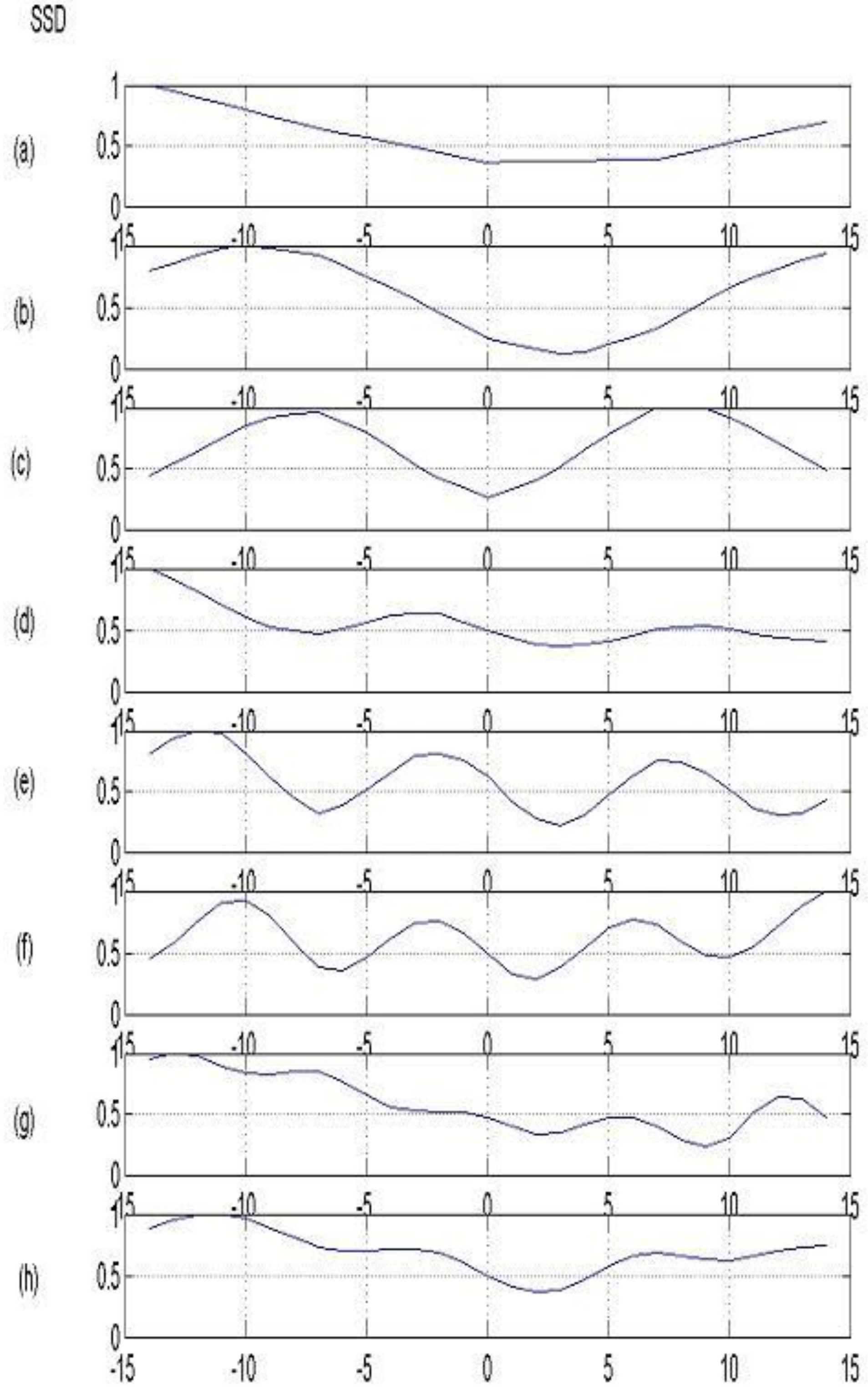
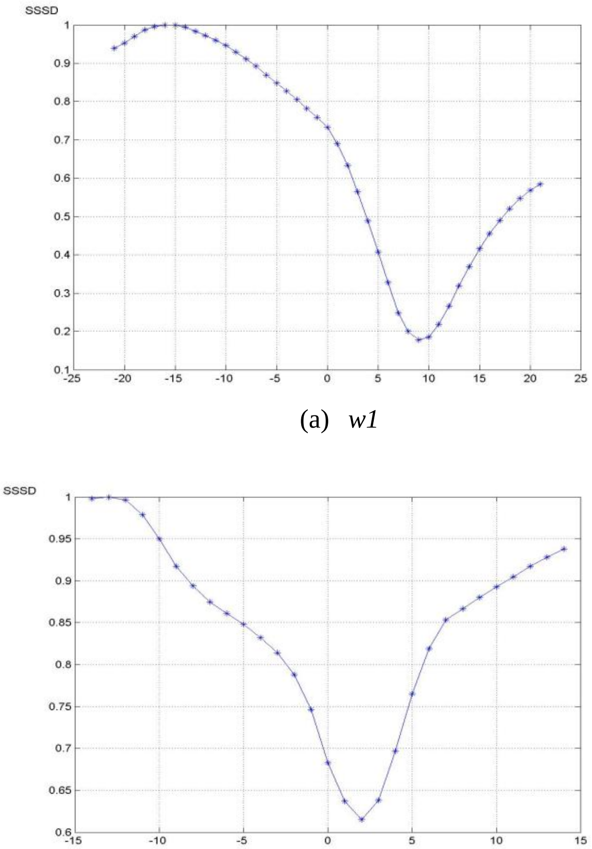


Figure 5.4: The evaluation functions versus distance. (Δ is the baseline, R is the searching range. The horizontal axis is normalized to *7b/ψF*=1. The vertical axis is normalized to [0~1]. (a) Δ=b; R=2 (b) Δ=2b; R=4 (c) Δ=3b; R=6 (d) Δ=4b; R=8 (e) Δ=5b; R=10 (f) Δ=6b; R=12 (g) Δ=7b; R=14 (h) result obtained from the modified multi-baseline algorithm)

### 5.1.3.2 Improving the depth estimation precision

To measure the thickness of the matchbox in the UII data, the two windows (*w1* and *w2*) were chosen again, see Figure 4.19. However, the SSSD(D) function is used instead of SSD(d) function as evaluation criterion to accumulate the information from multiple viewpoint images. Figure 5.5 shows the accumulated score functions obtained from the modified multi-baseline algorithm. A disparity of 9 can be found for *w1*, which gives a depth of 19.1*mm* for the object region. A disparity of 2 is obtained for *w2*, and this estimates the background depth to be 4.2*mm*. Therefore, the thickness of the matchbox can be calculated as 14.9mm. The actual matching thickness, as measured, is 15.6mm and therefore the error is 4.5%. This represents an improvement of 13% over the result obtained using traditional block matching algorithm.



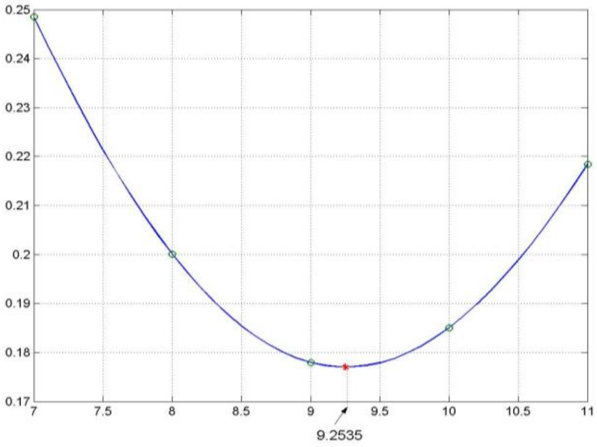
(a) *w1*

(b) *w2*

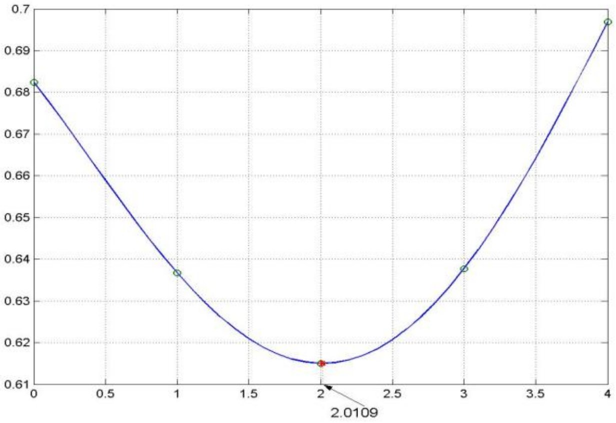
Figure 5.5: The score functions obtained by using multi-baseline technique.

(The horizontal axis is normalized to *7b/ψF*=1 and the vertical axis is normalized to [0~1])

The matching analysis only gives a discrete displacement. A more precise matching position can be obtained by using sub-pixel analysis. Based on the fact that the actual precise matching position is obtained from the actual minimum position on the score function, a polynomial curve was used to fit the function around the minimum position. The minimum of the polynomial curve can be regarded as the minimum position of the score function. Figure 5.6 shows the two polynomial curves obtained for the two minimum positions from Figure 5.5. The precise matching position can be found at 9.2535 for *w1* and 2.0109 for *w2* from the continuous functions. This gives the depth of the matchbox surface as 19.6*mm* and the background plane for 4.3*mm*, respectively. The estimated thickness of the matchbox is 15.3*mm* on the results. When compared to the manually measured value (15.6*mm*), the relative error is less than 2% in this measurement.



*(a) w1*



*(b) w2*

Figure 5.6: Precise solution finding from polynomial curve fitting.

Table 5.1 lists all the depth measurement results obtained for the matchbox using different algorithms. The improvement, using the modified multi-baseline with polynomial curve fitting, is obvious.

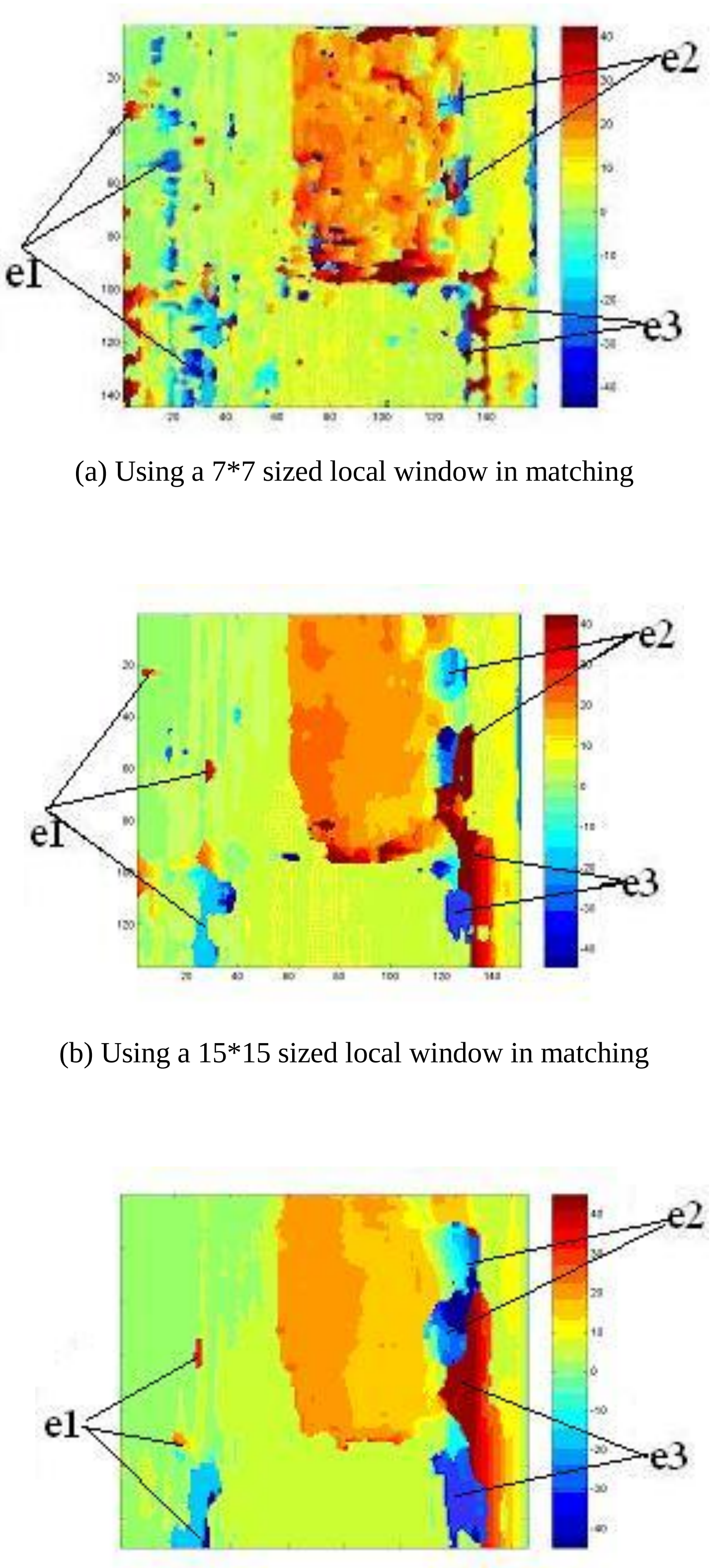
Table 5.1: Comparison of the depth measure results on matchbox using different algorithms based on SSD criterion

| Method used | Thickness of the matchbox (*mm*) | Error |
|---|---|---|
| Mean | 18.4 | 17.9% |
| Median | 17.3 | 10.9% |
| Multi-baseline | 14.9 | 4.5% |
| Multi-baseline with polynomial curve fitting | 15.3 | 1.9% |

### 5.1.3.3 Generating the depth map and reconstructing the 3D object

a) Photographically captured UII data

A dense depth map of the object scene can be obtained by applying the developed multi-baseline algorithm to every position of the viewpoint image. Generally, under the condition that all the pixels within the matching window are at the same depth, a larger local window is more noise-resistant than a small window in the depth extraction task since a large window contains more information and is easier to distinguish from the false matching positions. However, with the increase in window size, pixels at different depths start to enter the matching window and make correct matching difficult. To investigate the effect of window size, several different local window sizes are used for comparison. The corresponding depth maps are shown in Figure 5.7.



(a) Using a 7*7 sized local window in matching

(b) Using a 15*15 sized local window in matching

(c) Using a 21*21sized local window in matching

Figure5.7: The depth maps obtained for the matchbox with different matching window sizes.

For most regions, the depth is correctly estimated. The contour of the matchbox is clearly delineated on the three depth map generated from different matching window sizes. The errors can be mainly divided into three categories:

*e1*, error caused by lack of texture

*e2*, error caused by occlusion.

*e3*, error caused by illumination change in different areas, including shadows.

The error *e1* caused by lack of texture can be effectively reduced by increasing the window size used in matching. In contrast, the error *e2* caused by occlusion (around the object border) increases as the matching window size increase.

For both the object region and the background region, the colour in the depth map is slightly deeper on the left side. This indicates the object scene was not normal to the camera, as illustrated in Figure 5.8.

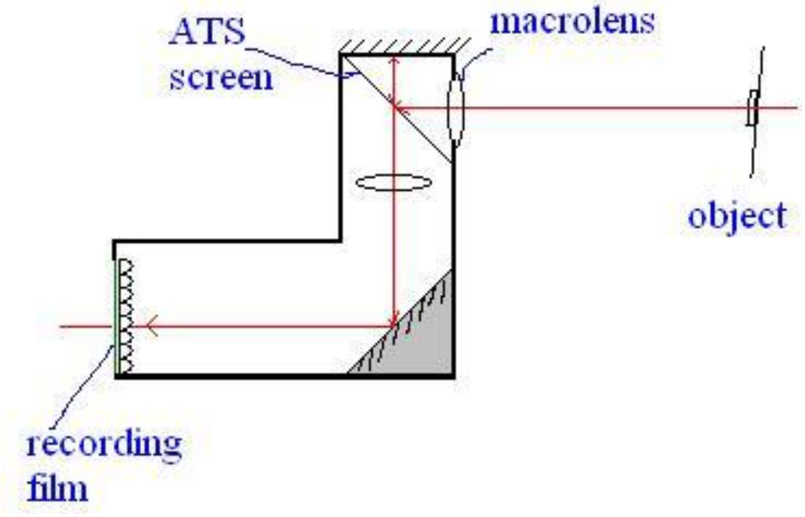


Figure 5.8: The object scene used in taken the UII data, matchbox.

Having obtained the depth, the 3D object space reconstruction is straight forward.   This is carried out by mapping the intensity information onto the depth results.  The reconstructed object scene for the matchbox is shown in Figure 5.9.

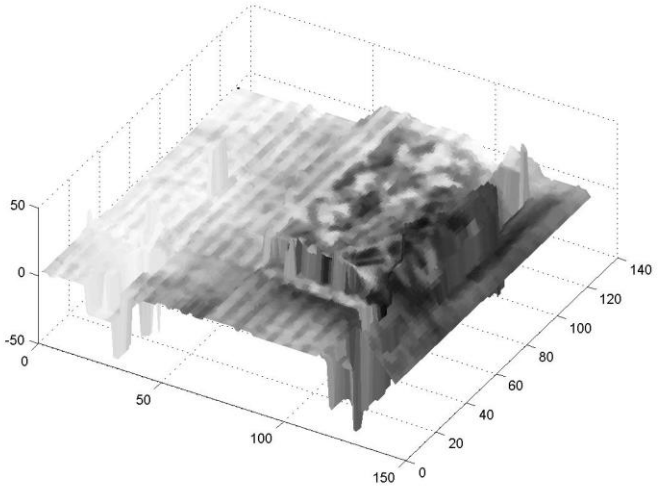


Figure 5.9: Reconstruction of the 3D object from the Captured UII data (matchbox) using the depth map generated from a 21*21 local window size.

b)        Computer generated UII data

To test the feasibility of the depth estimation approach on computer generated UII data, a synthetic UII data (cg_box), which contains the similar object scene as the captured UII data (matchbox) was computer generated using a modified PoV-Ray software developed by Cartwright in 3D Image Techniques groups, De Montfort University (Cartwright 2000).   The package uses a Ray-tracing technique in produce images in a way similar to the operation of the camera, casting rays out into the scene in a camera model based on pinhole model (Appel 1968, Povray 1999).  The modification is carried out by interfacing the integral imaging code into the POV-Ray source.

Figure 5.10 shows the computer-generated UII data (cg_box). The parameters of the recording micro-lenses sheet are the same as those used in recording the photographic

UII (matchbox). A 3-D scene can be replayed in a same way as replaying the matchbox, see Figure 4.16.

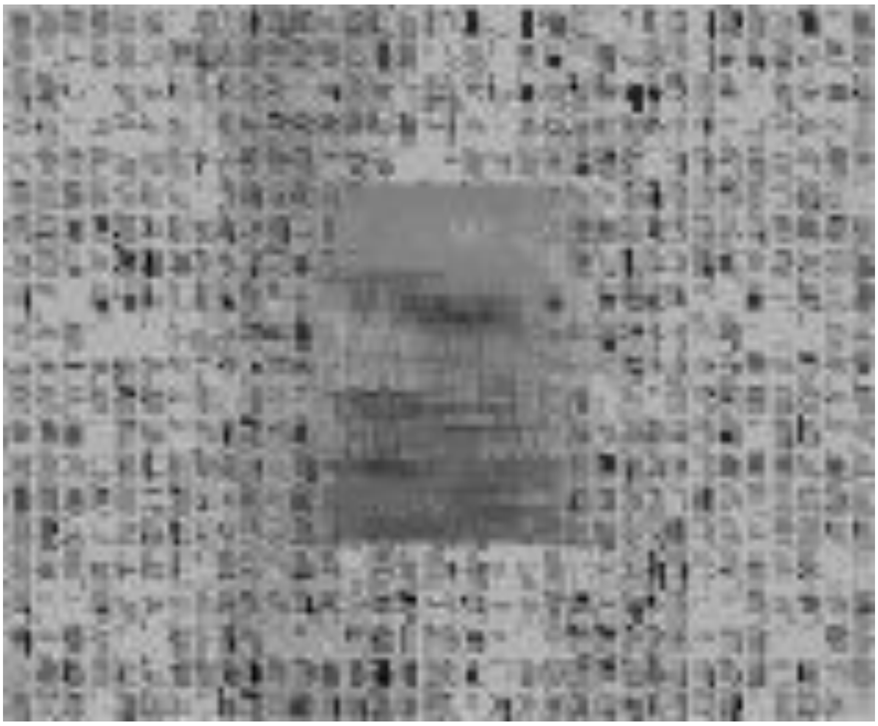
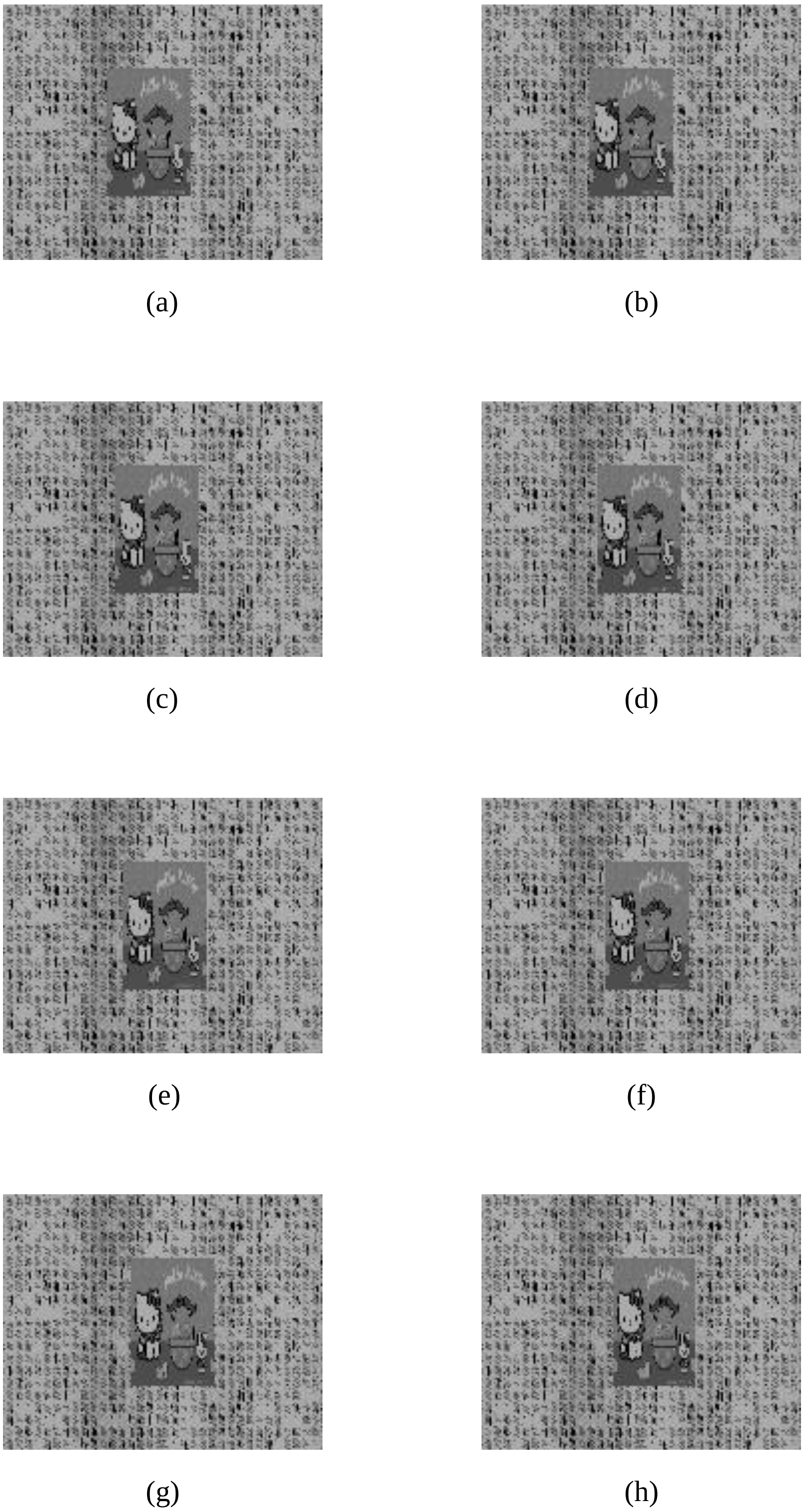


Figure 5.10: The computer generated UII data (cg_box). A 3D scene with a small box float abovet can be replayed as a 3D

Similar procedures were carried out to extract the depth information from the computer generated UII data. Figure5.11 shows the eight extracted viewpoint images and Figure 5.12 shows the depth maps obtained from different local window size. The reconstructed object space using the depth map obtained from the 7*7 matching window is shown in Figure 5.13.



(a)

(b)

(c)

(d)

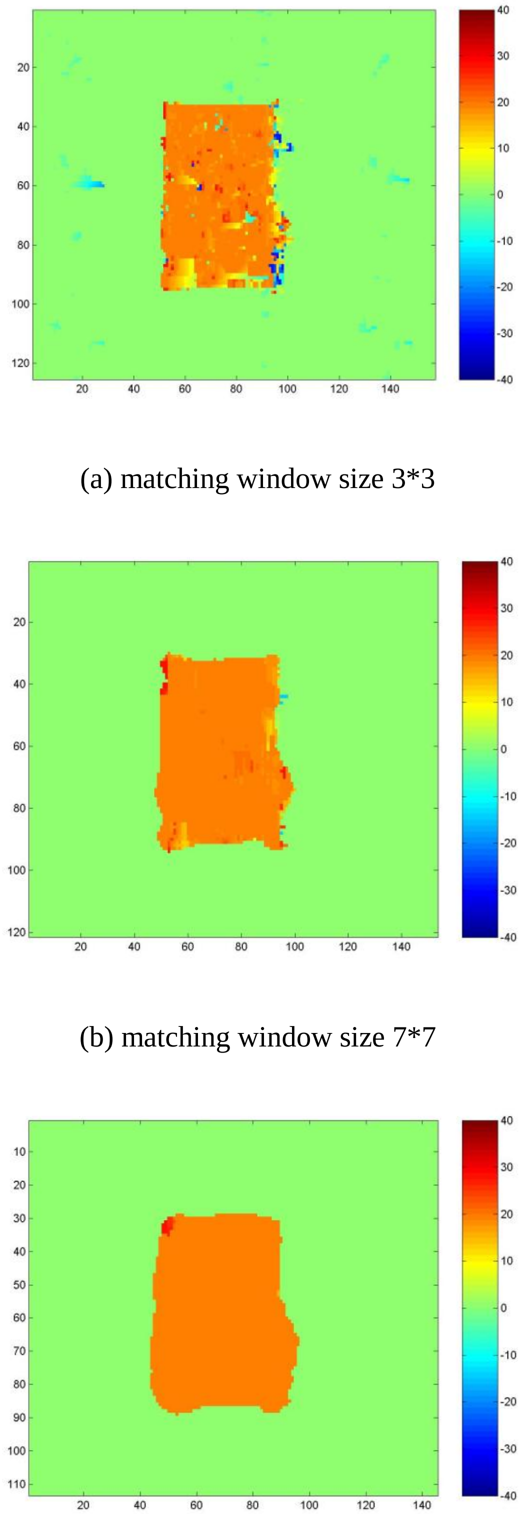
(e)

(f)

(g)

(h)

Figure 5.11: Eight viewpoint images extracted from the computer generated UII data (cg_box).



(a) matching window size 3*3

(b) matching window size 7*7

(c)  matching window size 15*15

Figure5.12: The depth maps obtained from the computer generated UII data (cg_box,)

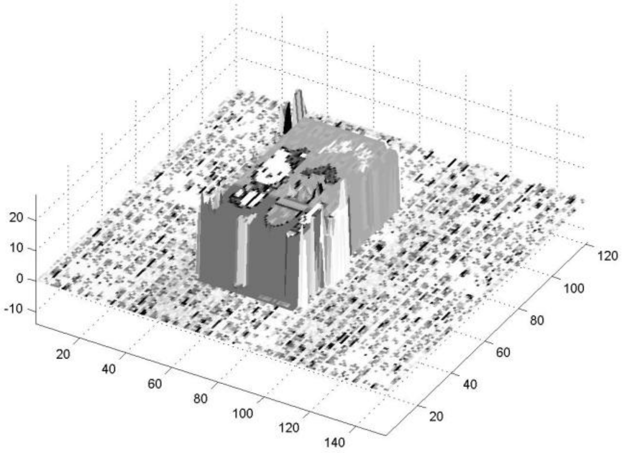


Figure 5.13: The reconstructed object space from the UII data (cg_box).

Generally speaking, a better resolution and depth estimation can be found from the computer-generated UII since the illumination is well controlled and also no noise is present in the recording process. The conclusion from choosing different window sizes is the same as before: a smaller matching window size gives more error within object/background region due to lack of enough features for correct matching. A bigger matching window size improves the matching results within object/background region but gives a worse contour of the object. To further improve performance, it was decided to explore the use of neighbourhood constraint and relaxation technique by considering the spatial constraint.

## 5.2 Disparity analysis using neighbourhood constraint and relaxation technique

This section presents an algorithm based on a neighbourhood constraint and relaxation technique to further improve the disparity analysis performance. Instead of using the simple SSD criterion as the matching evaluation function as in most literature, a relatively complicated criterion is used in considering the spatial constraint to determine the matching position. A multi-candidate pre-screen technique is also used to improve computation efficiency.

### 5.2.1 Theoretical foundation of the neighbourhood constraint and relaxation technique

1) Neighbourhood constraint

The neighbourhood constraint is based on the spatial-consistency rule. That is, the depth is piecewise continuous in the space (Orchard 1993, Dufaux 1995, Chen 1999). Therefore, the disparity can be more robustly estimated if the disparity within the neighborhood is considered. To better determine the matching position of a feature block, the neighbouring blocks ($N(B_{i,j})$) are considered rather than individually considering each block ($B_{i,j}$).

The neighbourhood of a block is illustrated in Figure 5.14.

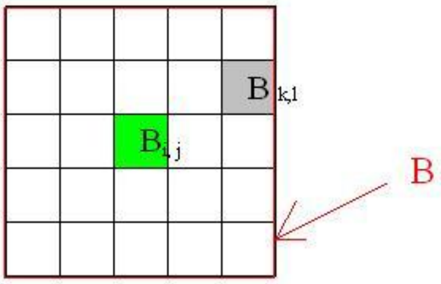


Figure 5.14: The neighourhood (B) of a block $B_{i,j}$ being analyzed.

If we consider the neighbourhood constraint, the new score function used in determining the matching position can be introduced as:

$$score(B_{i,j},\vec{d}) = SSD(B_{i,j},\vec{d}) + \sum_{B_{k,l}\in \mathrm{N}(B_{i,j})} W(B_{k,l},B_{i,j}) \;\; SSD(B_{k,l},\vec{d}) \qquad 5.13)$$

where, $B_{i,j}$ represents the window around pixel $(i,j)$, whose disparity is to be determined. $N(B_{i,j})$ is the set of neighbouring blocks of $B_{i,j}$, as shown in Figure 5.14. $w(B_{k,l},B_{i,j})$ is the weighting factor for the different neighbour blocks. The weighting scheme is introduced to reduce the estimation error caused when the neighbourhood block contains pixels in different depth by putting more emphasis on the center block than the periphery block. The weighting factors can be made to be dependent on several factors. For example, the distance to the central block, the confidence of the blocks and the color/texture similarity. In terms of the distance, a Gaussian-like function can be applied to the weighting factor based on the fact that objects are spatially continuous. The closer in spatial distance, the more possible they are having the same/similar depth. In terms of confidence, a block with large variance always contains more information for matching hence the result is more trustable. Therefore, it is always practical to put more emphasis on the blocks with larger variance than blocks with small variance. As for colour/texture similarity, based on the fact that the blocks within the same object usually have high texture/color similarity, it is reasonable to put high weighting factor on those blocks which have high similarity in color/texture characters to the centre block. By adjusting the weight factor, the neighbourhood of a block can be involved in determining the matching position by provide certain amount of support.

2) Neighbourhood relaxation

Neighborhood relaxation is used to allow for local variations of the disparity among neighbouring blocks in considering that the expected disparity of a neighbouring block is not necessarily equal to the centre block. To enable flexibility in considering the disparities of the neighbor blocks, a small $\vec{\delta}$ is incorporated to allow some disparity variations among the neighboring blocks.

Considering the neighbourhood relaxation, the score function can be written as:

$$score(B_{i,j},\vec{d}) = SSD(B_{i,j},\vec{d}) + \sum_{B_{k,l}\in N(B_{i,j})} W(B_{k,l},B_{i,j}) \min_{\vec{\delta}}\{SSD(B_{k,l},\vec{d}+\vec{\delta})\} \qquad 5.14)$$

This is the completed score criterion for neighbourhood constraint and relaxation. The neighbourhood constraint is implemented by summing the SSD functions of the windows in the neighbourhood. The neighbourhood relaxation is implemented by incorporating $\vec{\delta}$ to allow a certain degree of disparity vibrations among neighbouring blocks. The first item is treated as the image force which reflects the influence of the feature block and is similar to the external energy function of the popular SNAKE algorithm used in detecting object contour. The second item reflect the influence of the neighbours and is similar to the internal energy of the SNAKE algorithm.

The expected disparity is obtained on the position where the score function has the minimum.

$$\vec{d}^{*}{}_{i,j} = \arg\{\min_{d\in R}\{score(B_{i,j},\vec{d})\}\} \qquad 5.15)$$

3) Multi-candidate pre-screening

It is obvious that the Neighborhood Constraint and Relaxation criterion involves more calculation than traditional block matching considering the neighbourhood. To avoid unnecessary computation, a multi-candidate pre-screening strategy is used.

The idea is to carry out a two-stage voting scheme. This is based on the fact that the true disparity is likely to have a small residue but not always yields the minimum residue in the simple score criterion. However, implementation of a residue threshold by simple calculation the simple score criterion can at least preclude those positions that are unlikely to be the matching position. In the first stage, a simple criterion is used to choose candidates. In current case, the SSD function is used as the simple criterion. Final judgment is carried out in the second stage by using a more complete and complicated criterion on the chosen candidates.

The method of using multi-candidates pre-screening in obtaining the candidates can be described by the following steps:

I) Use the simple score criterion, SSD function to calculate the residue for all possible disparity positions. The minimum residue is obtained and recorded.

II) Choose a residue threshold ($R_{th}$): The residue threshold is always set as $\eta$ times of the minimum residue to allow a suitable number of candidates.

$$R_{th} = \eta * \min\_\,residue\,,\ \ (\eta > 1). \qquad 5.16)$$

III) Choose the candidates identified by the residue threshold ($R_{th}$). All searching positions having residues lower than the threshold are accepted as candidates while others are rejected.

Figure 5.15 graphically illustrates how the candidates are chosen according to the residue threshold. The larger the $R_{th}$, the more candidates are chosen and involved in the further competition, vice versa. After Pre-screening process, only those competent candidates are kept for further process using the complete evaluation function. This effectively reduces the computation.

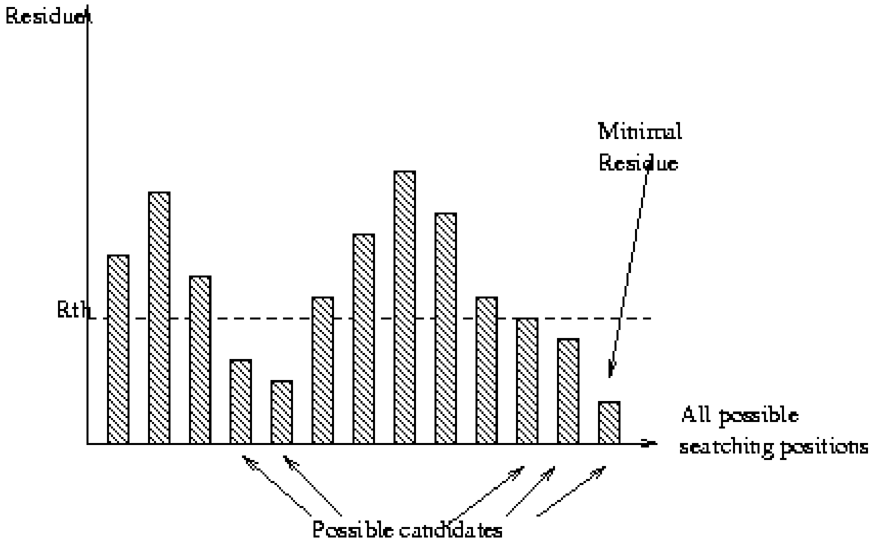


Figure 5.15: Multi-candidate pre-screening.

### 5.2.2 Implementation of the 'neighborhood constraint and relaxation' algorithm

The 'neighbourhood constraint and relaxation' algorithm is implemented using C language under the UNIX operating system. The algorithm mainly contains two modules:

1) Use multi-candidate pre-screening technique to choose candidates for matching position.
2) Use neighbourhood constraint and relaxation criterion to find the correct matching position from the candidates.

The main parameters used in the program are:

*1) Basic block matching window size (bw)*

*2) Residue threshold ($R_{th}$) parameter $\eta$*

*3) Neighbourhood block number (NBN)*

*4) Neighbourhood weight factor (NWF)*

Figure 5.16 shows the arrangement of the neighbourhood block sequence in the algorithm. The neighbourhood block is arranged in a sequence according to the distance to the centre block. The closer blocks (1,2,3,4) are followed by farther blocks (5,6,7,8). The number *NBN* defines the number of blocks in the sequence that can be chosen as 'neighbour'. A bigger *NBN* gives a bigger neighbourhood. Usually, NBN is chosen as the following numbers: 4, 8, 12, 20, 24, 28, 36, 44 and 48. As an example, when *NSN*=12, all blocks with number no more than 12 (marked in grey in figure 5.16), are chosen as the neighbourhood the feature block.

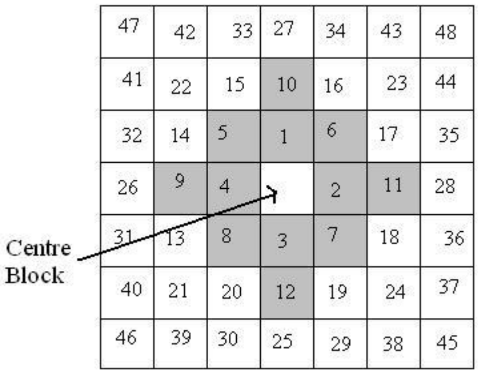


Figure 5.16: The neighbourhood blocks sequence.

In the algorithm, the NWF is defined according to the inverse distance to the central block, as shown in table 5.2. Hence, a close neighbouring block gives more contribution to the evaluation function. The further the distance between the neighbouring block and the center block, the less effect the block has on the final score function.

Table 5.2: The NWF for the neighbourhood blocks shown in Figure 5.16.

| Neighbouring block number | NWF |
|---|---|
| 1~4 | *nw* |
| 5~8 | 0.707 *nw* |
| 9~12 | 0.5 *nw* |
| 13~20 | 0.447 *nw* |
| 21~24 | 0.354 *nw* |
| 25~28 | 0.333 *nw* |
| 29~36 | 0.316 *nw* |
| 37~44 | 0.277 *nw* |
| 45~48 | 0.236 *nw* |

To combine with the multi-baseline technique used in previous chapter, a *SSSD(D)* function is used to replace *SSD(d)* function in the score function in equation 5.15). The score function combining two techniques can be written as:

$$score(B_{i,j},\vec{D}) = SSSD(B_{i,j},\vec{D}) + \sum_{B_{k,l}\in \mathrm{N}(B_{i,j})} W(B_{k,l},B_{i,j}) \min_{\vec{\delta}}\{SSSD(B_{k,l},\vec{D}+\vec{\delta})\} \qquad 5.17)$$

The output result is a data file that contains the corresponding disparity between viewpoint images.

Procedure 5.1 and 5.2 are the two major modules used in the programming:

### Procedure 5.1: *multi-candidate pre-screening*

```
Input arguments:
Thres              {Residue threshold ( $R_{th}$ )parameter η, ( $\eta > 1$ ) }
dfd.search_range   {the searching range for  two adjacent viewpoint image pair}
Results:
dfd.candidate      {mark the candidates}
Functions:
get_best_residue   {get the minima residue of the matching block on SSSD criterion }
get_residue        {get the residue for the searching position on SSSD criterion }

For each feature block
    minima_ residue= get_best_residue ()
    residue_threshold =Thres*minima_residue
    For each searching position  dfd.search_range
        residue=get_residue()
        If residue< residue_threshold
            Set the corresponding position of dfd.candidate to valid
        Else
             Set the corresponding position of dfd.candidates to invalid
```

### Procedure 5.2: N*eighbourhood constraint and relaxation*

```
Input arguments:
Thres              {Residue threshold ( $R_{th}$ )parameter η, ( $\eta > 1$ ) }
Results:
dfd.candidate      {mark the candidates}
Functions:
get_l_residue        {get the min_residue for the block on a displacement within δ
                     relaxation}
get_residue         {get the residue for the searching position on SSSD criterion }

For each feature block  ( $B_{i,j}$ )
    For each candidate searching position   $\underline{l}$
        Calculate the score:
            Set initial total_score=get_residue ( $B_{i,j}$ , $\underline{l}$ )
            Set initial valid =1;
            For each neighbouring block ( $B_{k,l}$ )
                total_score=total_score+w[ $B_{k,l}$ ]*get_l_residue( $B_{k,l}$ ,$\underline{l}$)
                total_valid=total_valid+w[ $B_{k,l}$ ]
     score=total_score/total_valid
record the min_score of score and its corresponding displacement vx
```

### 5.2.3 Experimental test

1) Captured UII data with simple object scene

The captured UII data, matchbox, is initially used to test the algorithm. Figure 5.17 shows the depth map obtained from the algorithm combining multi-baseline and neighbourhood relaxation & constraint technique using the initial parameters listed in Table 5.3. A reasonably good result can be obtained at the first instance except for the bottom left corner of the matchbox.

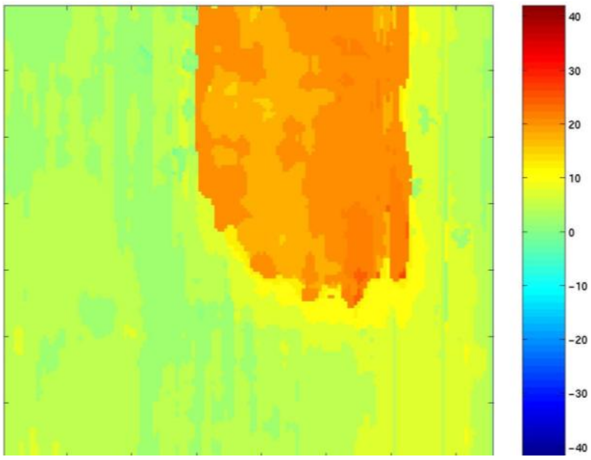


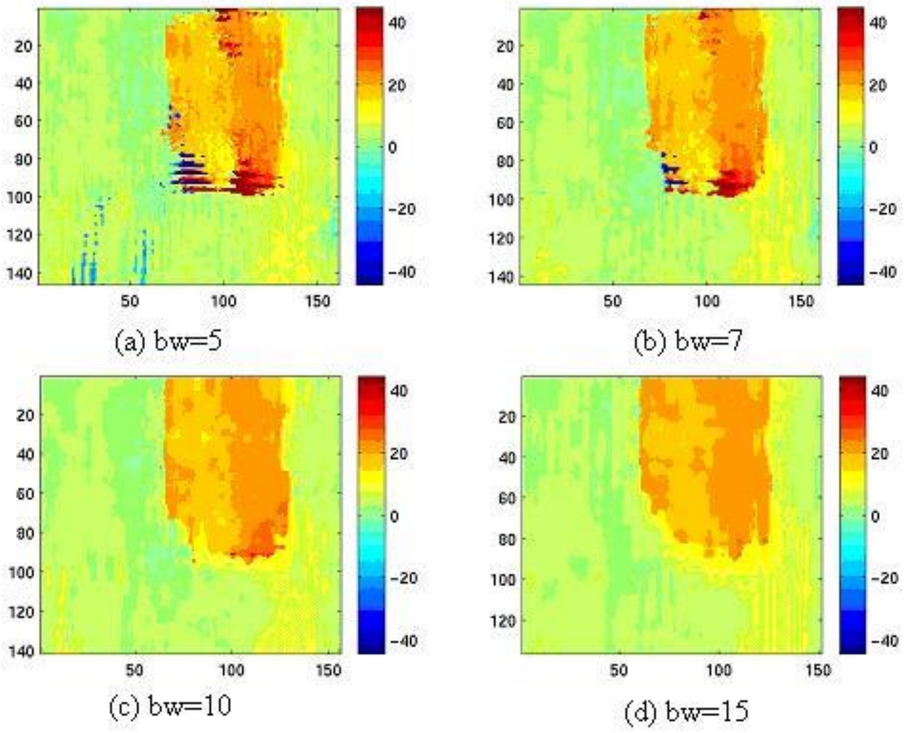
Figure 5.17: The depth map for the captured UII data, matchbox, using the initial parameters list in Table 5.3.

Table 5.3: The parameters used in generating Figure 5.17 from the UII data, matchbox.

| | Initial parameters | Final parameters |
|---|---|---|
| Valid viewpoint image numbers (N) | 7 | |
| Basic searching range (R) | 3 | |
| Basic Block Matching window size (*bw*) | 15 | 10 |
| Residue threshold ( $R_{th}$ ) parameter η | 2.0 | 2.0 |
| Neighbourhood block number (NBN) | 12 | 8 |
| Neighbourhood weight factor(NWF) | 0.5 | 0.5 |

The basic matching window size (*bw*) decides the basic measure scale. Figure 5.18 shows the depth maps obtained using different basic matching window sizes. Increasing the matching window size may increase the chance of false matching. However, the bigger the matching window size, the less the detail can be obtained on the depth map. In the current case, *bw* =10 is considered as a suitable parameter.

The neighbourhood block number (NBN) decides how many neigbhours are involved in calculation. A large neighbourhood block number means that more neighbourhood blocks are considered. Figure 5.19 shows the depth maps obtained using different neighbourhood block numbers. NBN=0 means no neighbourhood is involved in calculation. It can be seen from Figure 5.19 that an increase in neighbourhood block numbers leads to a smoother depth map. In the current example, NBN=8 is considered as a suitable parameter.



(a) bw=5

(b) bw=7

(c) bw=10

(d) bw=15

Figure 5.18: Comparison of depth maps obtained using different matching window sizes

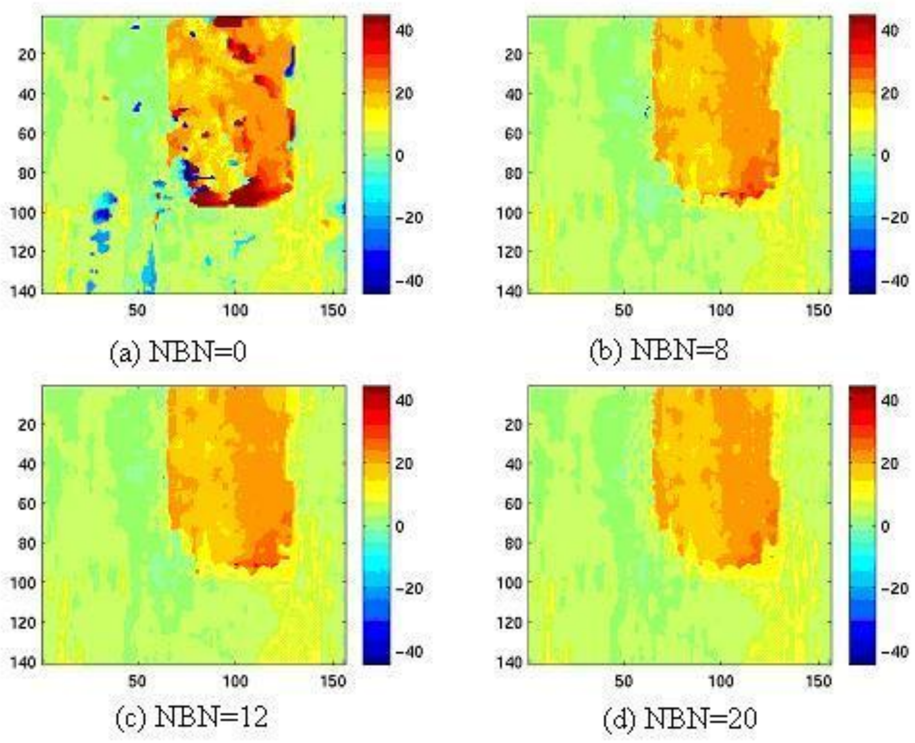
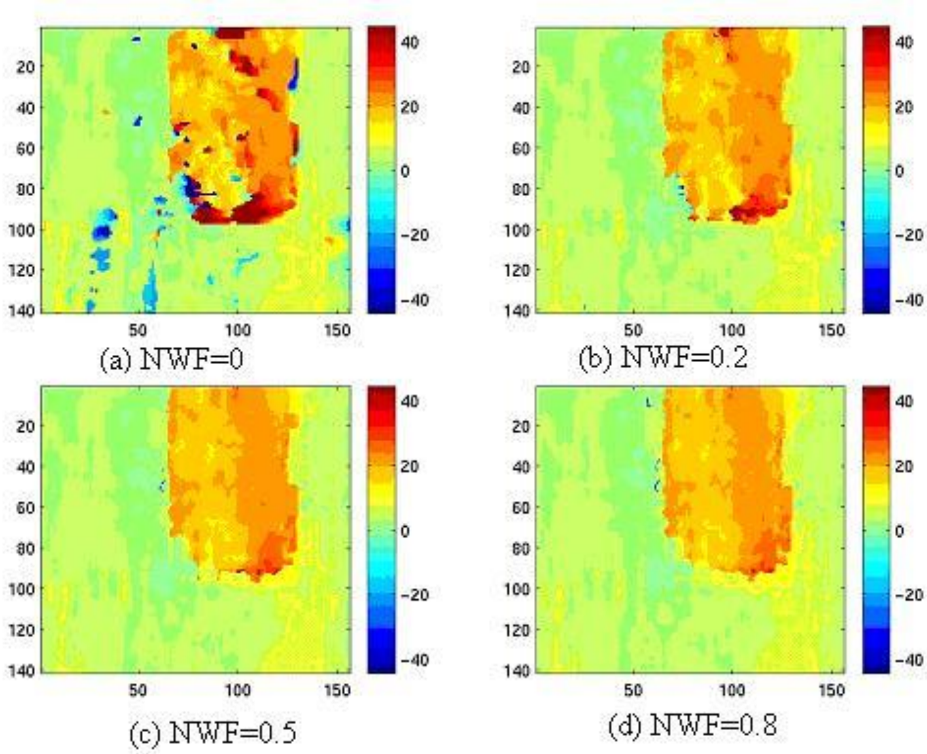


Figure 5.19: Comparison of depth maps obtained using different NBN (*bw*=10)

The Neighbourhood weight factor (NWF) support NBN in putting certain weight on the neighboring block. It is always set to a value between 0~1 so the centre block always has the principal influence. A high NWF enables more contribution from neighborhood blocks while NWF=0 means no contribution is received from the neighbouring blocks. Examining the results in Figure 5.20, NWF=0.2 is considered as a suitable parameter.

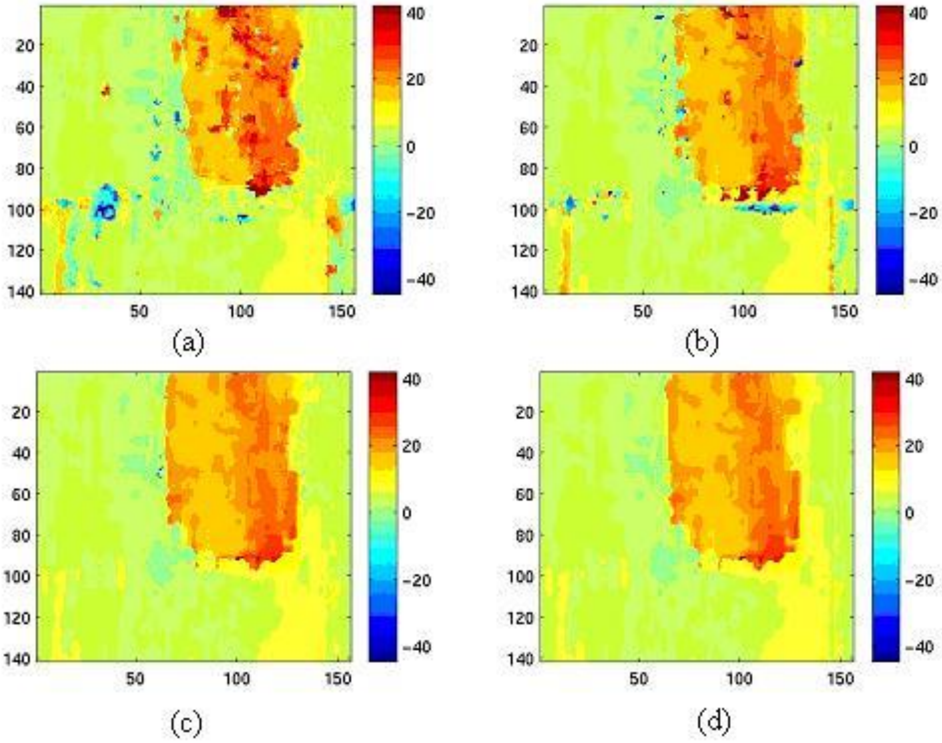
The residue threshold parameter $\eta$ acts as a factor in deciding the threshold for choosing candidates in the multi-candidates prescreening stage. A bigger $\eta$ will get more candidates and hence introduce more calculation. However, if the $\eta$ is set too small, the true matching position might be overlooked in the first round selection. As shown in figure 5.21(a) and (b), the false matching results are caused by setting a too small residue threshold parameter.



(a) NWF=0 (b) NWF=0.2

(c) NWF=0.5 (d) NWF=0.8

Figure 5.20: Comparison of depth map obtained using different NWF (bw=10, NBN=8)

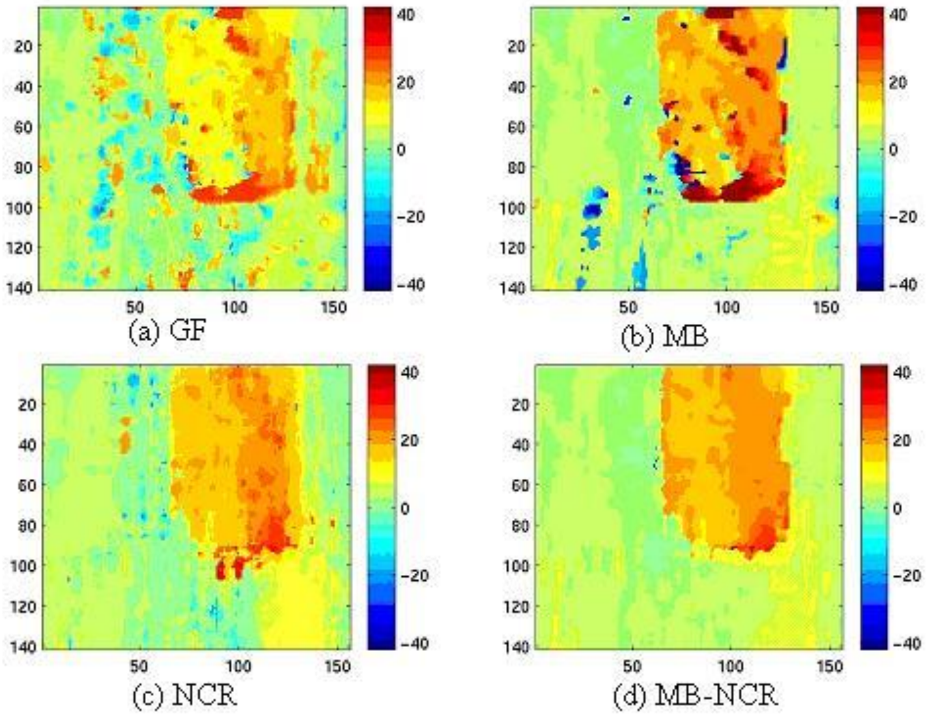


(a) (b)

(c) (d)

Figure 5.21: Comparison of depth map obtained using different Residue threshold parameters. (a) $\eta$=1.2 (b) $\eta$=1.5 (c) $\eta$=2.0 (d) $\eta$=5.0 (*bw*=10, NBN=12, NWF=0.5)

In the end, Figure 5.22 compares the depth maps obtained from four different algorithms: (a) GF algorithm: Use traditional block matching criterion, median of the results on different viewpoint image pairs is chosen as the final result. (b) MB algorithm: Use traditional block matching criterion with multi-baseline technique. (c) NRC algorithm: Use neighbourhood relaxation and constraint criterion and median for integrating the results from different pairs. (d)MB-NCR algorithm: Use neighbourhood relaxation and constraint criterion with multi-baseline technique. It is seen that the best result is achieved by using the MB-NCR algorithm. Compared with the depth map obtained from MB algorithm, almost all the false matching results exist in the object and background region has been removed with a good object contour perceived.



(a) GF

(b) MB

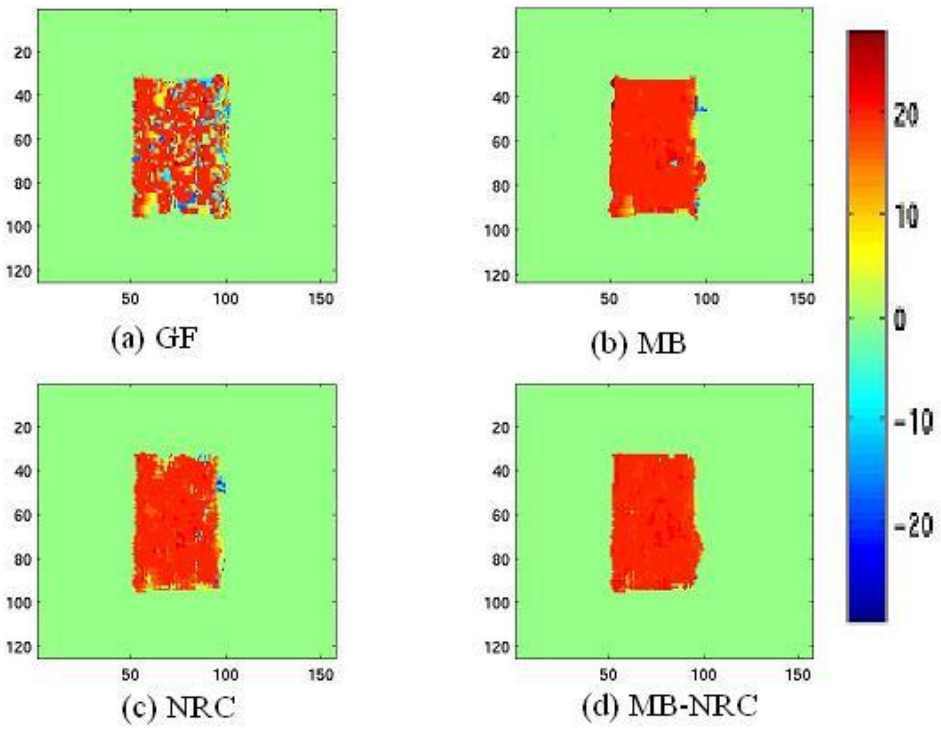
(c) NCR

(d) MB-NCR

Figure 5.22: Comparison of depth maps obtained using different algorithms on the captured UII data (Matchbox, *bw*=10).

2) Computer generated UII data with simple object scene

Figure 5.23 compares the depth maps obtained from the computer generated UII data (Cg_box), using the four different algorithms with parameters list in Table 5.4. Again, it can be notice that almost all the false matching results has been removed and a good object contour has achieved on the depth map using the MB-NRC algorithm.



(a) GF

(b) MB

(c) NRC

(d) MB-NRC

Figure 5.23: Compare the depth maps obtained from different algorithms on the computer generated UII data, cg_box4 (*bw*=3).

Table 5.4: The basic parameters used in generating depth maps for cg_box

| Algorithm | GF | MB | NRC | MB&NRC |
|---|---|---|---|---|
| Viewpoint image numbers (N) | 8 | | | |
| Basic searching range (R) | 3 | | | |
| Block Matching window size (*w*) | 3 | | | |
| Residue threshold parameter η | 5.0 | | | |
| Neighbourhood block number (NBN) | none | | 8 | |
| Neighbourhood weight factor | none | | 0.8 | |

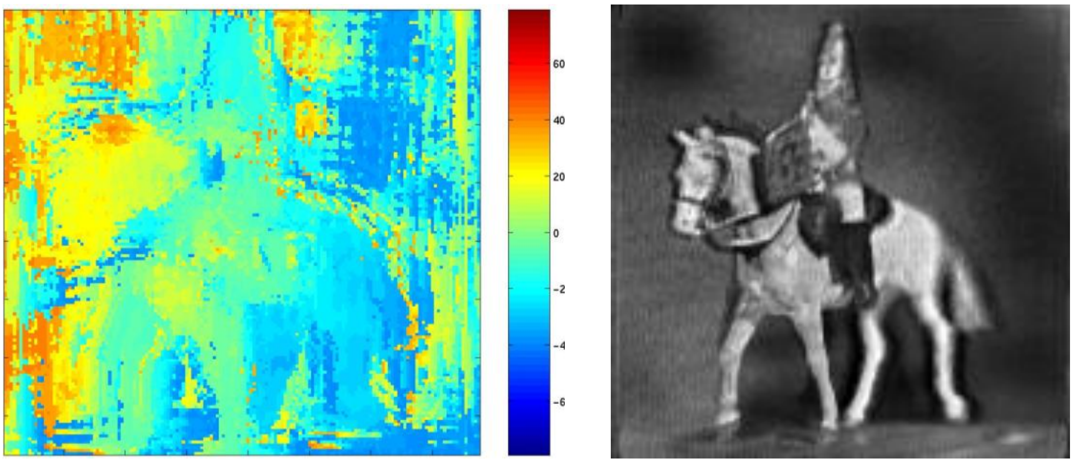
3) More captured UII data

Further experiments were carried out on more Captured UII data taken by researchers within the 3d Image Group. These images are shown in Annex, as Image A.2 (Horseman), Image A.3 (Tank) and Image A.4 (Lab). All the images can be replayed by using a microlens sheet with suitable parameters. Figure 5.24(a) ~ 5.26(a) shows the extracted depth maps. For comparison, the 2D views of each scene are shown in figure 5.24(b) ~ 5.26(b), respectively. The parameters used in the algorithm in generating the depth map are listed in table 5.5.

I) For figure 5.24(a), the different depths of the horse and the man who sits up on the horse can be correctly perceived from the extracted depth map. The head of the horse appears in the colour yellow. The forelegs of the horse appear in the colour green. The man is in the colour light blue and a slightly darker blue colour was obtained for the rear-legs and tail of the horse. The recording plane is found out to be positioned around the forelegs of the horse in the object scene. Poor results are obtained on the background region due to the lack of enough features in the region for matching.

II) For the second image (Tank), it is difficult to distinguish the contour of the tank from Figure 5.25(a). However, the colour variation gives a blur perception of the base under the tank. A brighter color means a position in the front of the scene. Again, poor results are obtained in the background region.

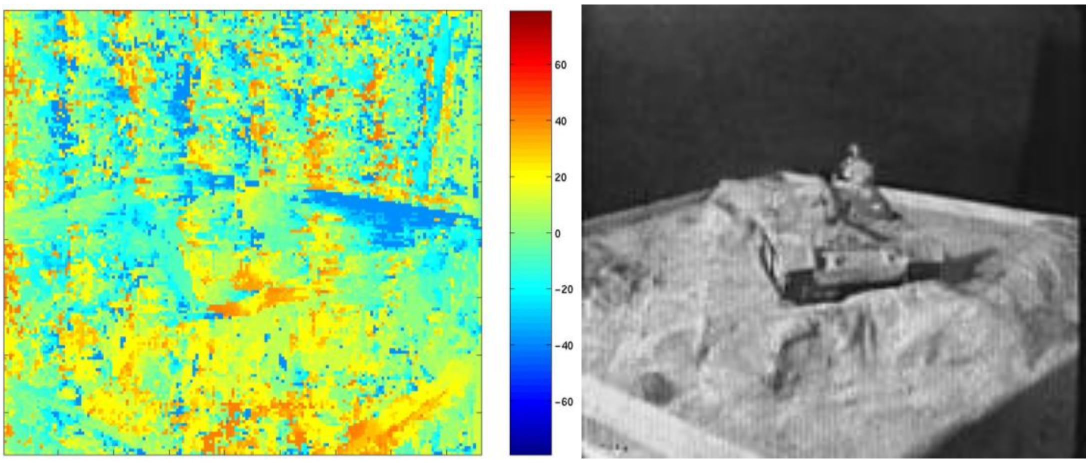
III) The last image (Lab) contains a quite complicated object scene with multiple objects and a natural background of a laboratory. In the middle of the object scene is a toy-plane placed in the foreground supported by a stand-frame behind it. The toy-plane is followed by a lamp on the left side and a man on the right. Some flowers are put in front of the man (at a position relative to the lower part of face and neck). The background is a half-open door with some flowers inside the door. On the depth map, Figure 5.26(a), most of the objects in the scene are perceived with correct depth. The toy-plane is in front of the scene plane in the colour orange followed with the supporting stand-frame in yellow. The contour of the lamp on the left side and the man on the right side can be just distinguished. The half open door in the rear of the scene appears in the colour blue.



(a) depth map

(b) 2D view image

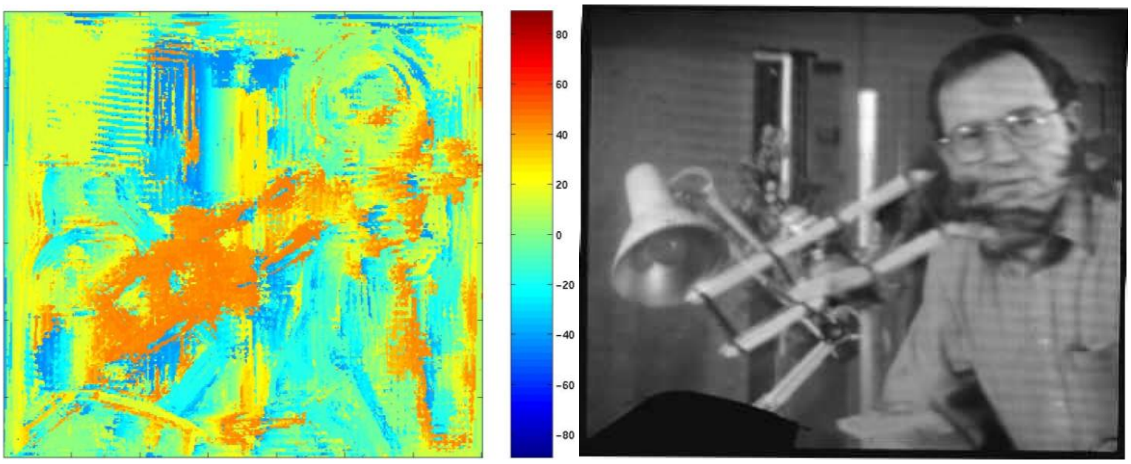
Figure 5.24: The depth map and one of the 2D views of the UII data (Horseman).



(a) depth map

(b) 2D view image

Figure 5.25: The depth map and one of the 2D views of the UII data (Tank).



(a) depth map

(b) 2D view image

Figure 5.26: The depth map and one of the 2D views of the UII data (Lab).

Table 5.5: The parameters used in obtaining Figure 5.24 ~ 5.26

| | Horseman | Tank | Lab |
|---|---|---|---|
| Viewpoint image numbers (N) | 8 | 8 | 12 |
| Valid viewpoint image numbers (VN) | 4 | 5 | 8 |
| Basic searching range (R) | 4 | 4 | 4 |
| Block Matching window size (*w*) | 5 | 4 | 5 |
| Residue threshold ( $R_{th}$ ) parameter η | 2 | 1.2 | 1.5 |
| Neighbourhood shape number (NSN) | 8 | 4 | 12 |
| Neighbourhood weight factor (NWF) | 0.8 | 0.8 | 0.8 |

# 5.3 Summary

This chapter presents the work using a modified multi-baseline algorithm with neighbourhood relaxation and constraint technique to improve the performance for the disparity analysis on the extracted viewpoint image. The first section concentrates on adapting the modified multi-baseline algorithm by using the information from multiple viewpoint images. Mathematical analysis on how the modified algorithm can remove ambiguities in the depth estimation from multiple viewpoint image pairs is given. Experiments proved the effectiveness of the algorithm. The depth measurement obtained from the matchbox using the modified multi-baseline algorithm with a polynomial curve fitting gives an error of less than 2% in the example. Depth map and object space reconstruction can be achieved from both the captured and computer generated UII data with acceptable quality. The size of the local window used in matching is found to be a factor that affects the final depth estimation result.

To further improve the performance of the disparity analysis, the second section presents a neighbourhood relaxation and constraint technique in considering the spatial constraint. This is carried out by using a more complicated score function in deciding the matching position through considering the spatial constraint from the neighbourhood. A multi-candidates prescreen technique is also used to improve the computation efficiency. Experiments have shown an obvious improvement on the depth map achieved from the MB-NCR algorithm combining both the neighbourhood relaxation and constraint criterion and multi-baseline technique. Several captured UII's contain complex object scenes taken by the researchers within this group are used. Using the MB-NCR algorithm combine two techniques presented in this chapter, most of the objects in the scene can be perceived at the correct depth position. The most prominent errors are found on the background region due to lack of enough features for matching.

# Chapter 6

# “Feature Block Pre-selection” and “Consistency Post-screening”

Following the work in previous chapters, this chapter examines approaches to improving the precision and correctness of the results gained from the disparity analysis of viewpoint images. As reported in chapter 5, it is within the homogenous regions that errors frequently appear due to the lack of sufficient features within the matching window. Another difficult region is around object borders where two different displacements exist within the matching block. In this chapter, two techniques, “feature block pre-selection” and “consistency post-screening” are employed to deal with these two situations. Experiments have shown that the two techniques worked successful in detecting those results with low confidence. Combined with the techniques used in previous chapters, it is anticipated that improved results can be achieved for the depth map generated from the UII data.

# 6.1 Feature block pre-selection

1) Untraceable block

In chapter 4, the basic idea of the block-matching method to locate a candidate block in the second image that can best match a target block in the first image is described. This is carried out by comparing the residues between two blocks. Matching is found at the position where a minimum residue is obtained. The confidence of matching depends on the information contained within the block. The more information contained within the matching window, the more noise-resistant is the result. However, information details are not uniformly distributed in the spatial domain, that is, some parts of the image have more details than others. One extreme circumstance is a block without texture, that is where the intensity is constant within the block. The block of pixels may look identical to another block of pixels and consequently more than one position will give the minimal-residue. In this case, the minimal-residue criterion does not always deliver the true matching position for depth estimation. We call these blocks “untraceable”. Another case might be a block that does not contain prominent texture features and finding the correct matching position is extremely difficult due to image noise disturbance. Consequently, the disparity results obtained for these blocks will have low confidence for depth estimation.

2) Detecting the untraceable blocks

The untraceable block causes false matching and thereby a faulty depth analysis process. For a good matching efficiency, it is advisable to identify these blocks before the matching process. A scheme to do this can be implemented by evaluating the variance of the blocks. When the intensity variance within the block is smaller than a given threshold, the block can be considered as “untraceable”. The matching results obtained from the “untraceable” blocks can then be recognized as having low confidence. Only the blocks that contain enough features for a confident match are kept for further analysis. This represents “feature blocks pre-selection” and is an efficient and informative procedure. In the final depth map, the depth of these untraceable positions can be recovered from the neighbourhood blocks and therefore the approach leads to an effective gain in the process.

## 6.2 Consistency post-screening

1) Untrackable block

The matching algorithm in disparity analysis relies on the assumption that the image intensity, corresponding to a point, remains constant in different images. However, it is not always true on all occasions. For example, the pixel intensity in the circle shown in Figure 6.1 is not constant due to occlusion.

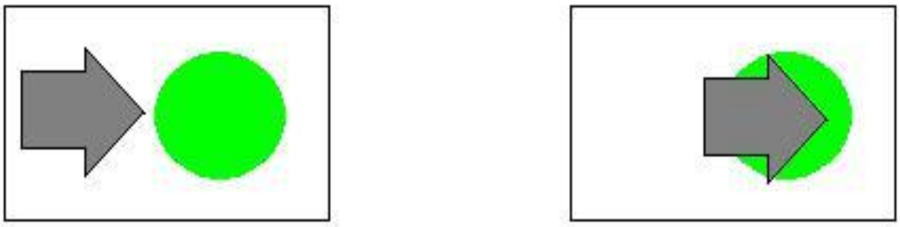


Figure 6.1 Occlusions occurs when viewing from different positions.

Other situations that can occur are where a block is located at the object boundary. Since at least two different displacements exist within the block, it is impossible to find a single disparity correctly to compensate the block. We call the object boundary, occlusion, and reappearance regions as “untrackable” regions. The exact matching position does not exist for “untrackable” regions. It is extremely useful to identify these blocks in the matching process.

2) Detecting the untrackable block

The consistency post-screening technique is used to identify these untrackable blocks. This is carried out by evaluating the residue from the score function based on the following observations:

I) When the estimated disparity ( $\vec{d}^*_{i,j}$ )is close to the true matching position for block ($B_{i,j}$) and $\vec{d}^*_{k,l}$ is close to the true matching position for block($B_{k,l}$) , the corresponding residue should satisfy:

$$SSD(B_{i,j},\vec{d}^*_{i,j}) < R_{th}(B_{i,j}) \text{ and } SSD(B_{k,l},\vec{d}^*_{k,l}) < R_{th}(B_{k,l}) \qquad 6.1)$$

II) If block $B_{k,l}$ and $B_{i,j}$ are in a same object, $\vec{d}^*_{i,j}$ and $\vec{d}^*_{k,l}$ suppose to close to each other, $\vec{d}^*_{k,l} = \vec{d}_{i,j}^{\,*} + \vec{\delta}$ and the residue still can satisfy:

$$SSD(B_{k,l},\vec{d}^*_{i,j}+\delta) < R_{th}(B_{k,l}) \qquad 6.2)$$

In total, the residue from the score function after considering the neighbourhood will be less than the weighted sum of the residue thresholds:

$$\begin{aligned} &SSD(B_{i,j},\vec{d}^*_{i,j}) + \sum_{B_{k,l}\in N(B_{i,j})} w(B_{k,l},B_{i,j})\cdot \min_{\delta}\{SSD(B_{k,l},\vec{d}^*_{i,j}+\delta)\} \\ &< R_{th}(B_{i,j}) + \sum_{B_{k,l}\in N(B_{i,j})} w(B_{k,l},B_{i,j})\cdot R_{th}(B_{k,l}) \end{aligned} \qquad 6.3)$$

III) If block $B_{k,l}$ and $B_{i,j}$ are in different objects in different depths, $\vec{d}^*_{i,j}$ is away from $\vec{d}^*_{k,l}$, equation 6.2) no longer holds. In most situations:

$$SSD(B_{k,l},\vec{d}^*_{i,j}+\delta) >> R_{th}(B_{k,l}) \qquad 6.4)$$

Therefore equation 6.3) is unlikely to be satisfied.

Consistency post-screening is carried out by evaluating all results with the estimated disparities using the relationship given in equation 6.3). The blocks that fail in the criterion are recognised as untrackable and can be marked out and discarded. The depth of these positions is later recovered by considering the neighbouring.

# 6.3 Implementation of the hybrid algorithm

The hybrid algorithm combining “Feature block pre-selection” and “consistency post-screening” technique along with the previous multi-baseline and neighbourhood constraint and relaxation techniques is implemented using C language under the UNIX operating system. The algorithm contains five major modules:

1) Using feature block pre-selection to identify the untraceable blocks.
2) Using multi-candidate pre-screening to choose candidates.
3) Using neighbourhood constraint and relaxation criterion with multi-baseline technique to find out the correct matching position.
4) Using consistency post-screening to detect the untrackable blocks.
5) Recovering the invalid results by considering neighbouring positions.

In addition to the parameters used in the previous algorithm, the feature block pre-selection threshold parameter ($FB_{th}$), is contained in the hybrid algorithm. This parameter acts in the same manner as a factor in evaluating the requirement for the variance within the block for a confident matching estimation. A larger $FB_{th}$ requires more feature information for the matching window hence increasing the $FB_{th}$ gives more untraceable positions.

The residue threshold parameter $\eta$ has two functions in this algorithm. The first is to act in the same manner as in previous chapter in deciding the threshold for choosing candidates in the multi-candidate prescreening stage. A bigger $\eta$ will get more candidates. Another function introduced in the hybrid algorithm is to act as a threshold in evaluating the validity of the matching results in the consistency post-screening stage. A bigger $\eta$ allows a loose constraint in the post-screen evaluation stage and hereby less untrackable positions will be detected.

Two different schemes are used to recover the untraceable and untrackable regions. In order to achieve a good boundary for the objects, the median of the eight valid neighbouring points is used for the untrackable region, while the mean value with a weighting factor is used to compensate the untraceable region in order to have a good continuity within the object. The output is a depth map data file corresponding to the UII data.

Procedure 6.1~6.3 lists the three major program modules in addition to previous MB-NRC algorithm:

<u>Procedure 6.1: *feature blocks pre-selection*</u>

```
Input arguments:
*currFrame*     {intensity distribution for the first viewpoint image}
*var_thres*     { Feature block pre-selection threshold parameter ($FB_{th}$):}
Results:
*dfd.mode*     {mark the valid feature blocks}

For each blocks
    Calculate the mean intensity value within the block.
     For each pixel position
         Calculate the mean intensity value within the block
     Calculate the variance within the block
     If variance >  *var_thres*
        Set the corresponding position of *dfd.mode* to valid
     else
        Set the corresponding position of *dfd.mode* to invalid
```

<u>Procedure 6.2: *Consistency post-screen*</u>

```
Input arguments:
*Frames_vx*              {the estimated disparity results from neighbourhood
                             constraint and relaxation analysis }
Results:
*Frames_vx*              {mark the untrackable disparity results}
Functions:
*get_residue*          {get the residue for the searching position using SSD criterion }
*get_residue_th*        { get the residue threshold of the block }

For each valid feature block ( $B_{i,j}$ )
     For each valid neighbouring block  ( $B_{k,l}$ )
```

```
            Find the minimum residue of the block ($B_{k,l}$) around the estimated
             displacement within a certain range vibration.
            Accumulate the residue with weight  *(score)*
            Accumulate the threshold with weight *(score_th)*
        If (score>score_th)
            Set the corresponding results to untrackable
```

Procedure 6.3: *Postprocess* *(recover the invalid depth results from its neighborhood)*

```
    Input arguments:
    Frames_fmv            {integrated disparity results}
    Results:
    Frames_fmv

    For each disparity result (i,j)
        If (Frames_fmv(i,j)=untraceable)
            Frames_fmv(i,j)=mean(the eight adjacent valid results)
        If (Frames_fmv(i,j)=untrackable)
            Frames_fmv(i,j)=median(the eight adjacent valid results)
```

# 6.4 Experiments

1) Computer generated UII data

The hybrid algorithm was first applied to the computer generated UII data, cg_box, to test for the feasibility. Figure 6.2 shows the depth maps obtained from the hybrid algorithm using different feature block pre-selection threshold parameters ($FB_{th}$). To illustrate the effects of using feature block pre-selection and consistency post-screening techniques, the untraceable positions are marked in dark blue while the untrackable positions are marked in dark red in the depth maps. The other parameters used are the same as in the previous chapter, see Table 5.4.

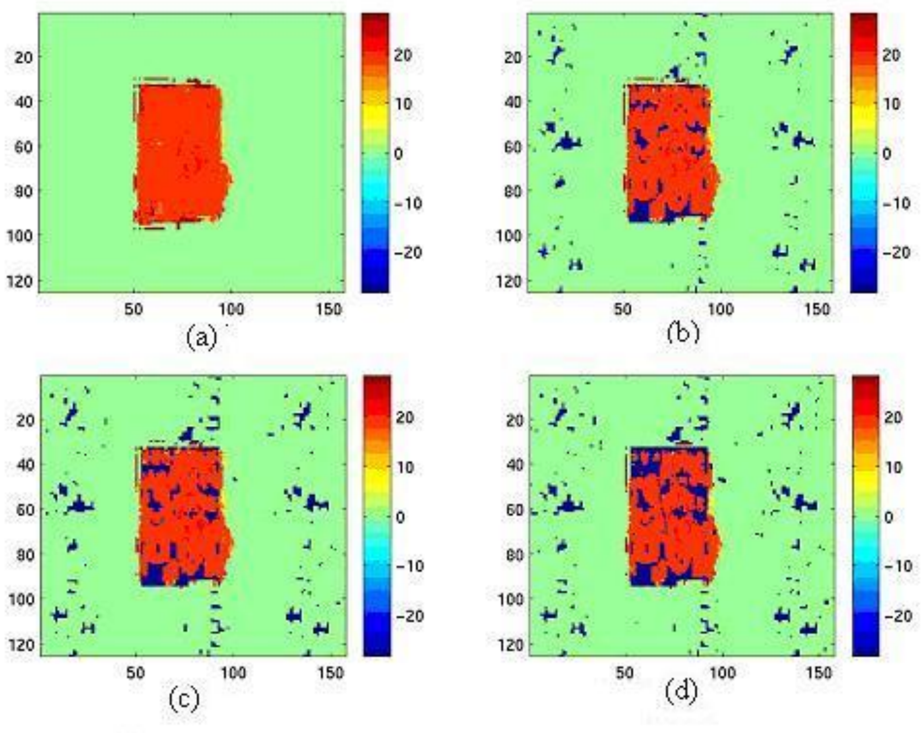


Figure 6.2: Comparison of depth maps obtained using different feature block prescreen selection threshold: (a)$FB_{th}$=0, (b)$FB_{th}$=1, (c)$FB_{th}$=2, (d)$FB_{th}$=3

More untraceable positions are detected when the feature block prescreen selection threshold is increased. The positions detected as untraceable correspond to the low intensity variation regions in the object scene, see Figure 5.11(a). As expected, the detected untrackable positions appear around object borders. In the image being analysed, only very small intensity variance ($FB_{th}$>0) within the matching window can lead to correct depth estimation. This is because the simple UII data used in the example is particularly designed for the matching task and no recording noise existing in the computer generated UII data.

A further test was carried out on another computer generated integral image (cg_balls) which contains several objects at different depths and a blank background. The UII data is shown in Annex, Image A.6. Figure 6.3 shows one of the 2D viewpoint images in grey level coding. The object scene contains a red ball in the middle of the scene followed by a golden ribbon-like object underneath. In the upper-left part of the scene is a blue ball behind the recording plane. A yellow ring is positioned the rear most.

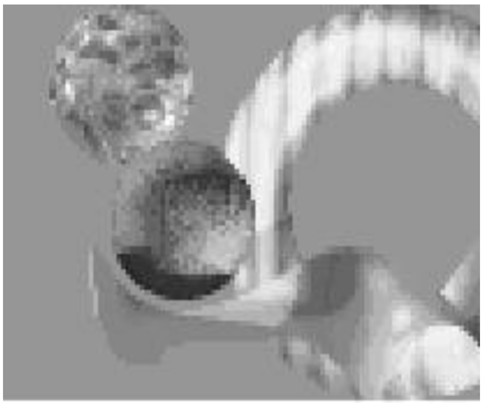


Figure 6.3: View images from cg_balls in gray level coding.

Figure 6.4 shows the corresponding depth map obtained from the hybrid algorithm after recovering the invalid results by its neighbours. The parameters used are listed in Table 6.1. Objects at different depths have been successfully detected: The red ball in the foreground appears as the colour orange. The golden ribbon-like object that traverses the scene and passes beneath the red ball is colour coded from orange through yellow to green. The blue ball behind the recording plane appears in the colour light blue and the yellow ring in the rear-most is blue. The background region (without pattern) is detected as untraceable and is represented by dark blue. The positions of the objects can be clearly interpreted from the depth map given in Figure 6.4. Continuous depth variation is perceivable. For example, the depth variation of the red ball can be clearly appreciated from the colour coded depth map.

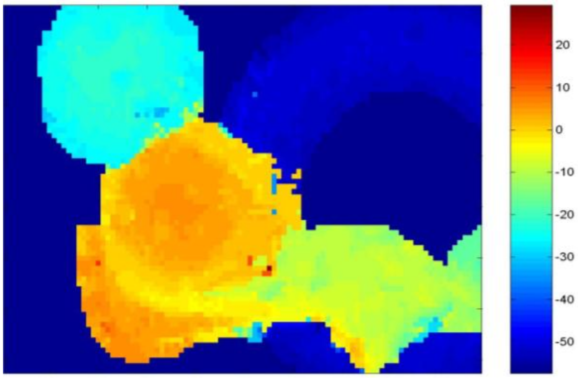


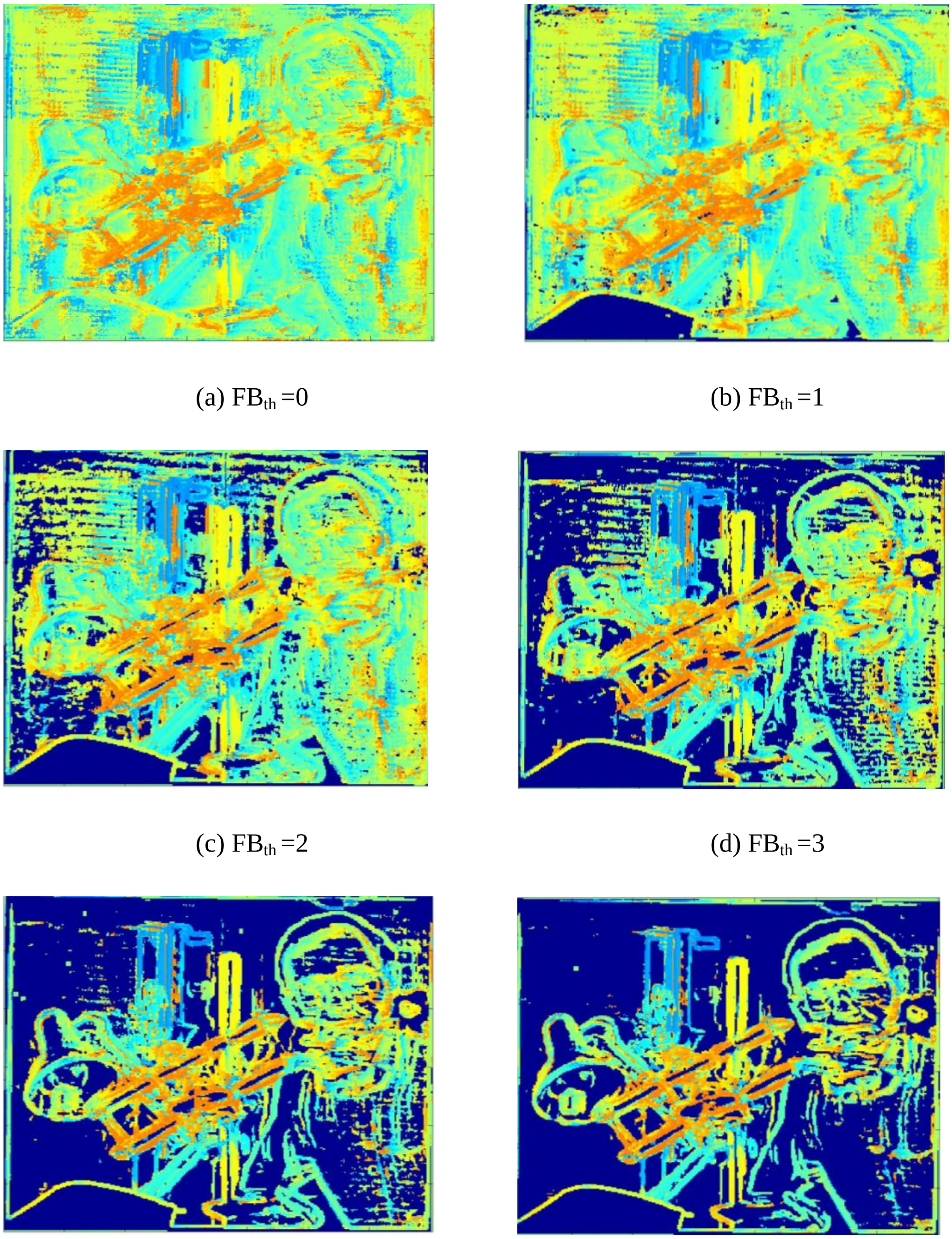
Figure 6.4: The depth map for cg_balls

Table 6.1: The parameters used in obtaining Figure 6.4

| | |
|---|---|
| Viewpoint image numbers (N) | 8 |
| Basic searching range (R) | 6 |
| Block Matching window size (*w*) | 3 |
| Feature block pre-selection threshold parameter ($FB_{th}$) | 1 |
| Residue threshold ( $R_{th}$ ) parameter η | 2.0 |
| Neighbourhood shape number (NSN) | 12 |
| Neighbourhood weight factor | 0.8 |

2) Captured UII data

After successfully applying the algorithm to computer generated integral images; a further test was carried out on captured photographic integral images. Initially, the UII data (Lab), which contains a quite complicated object scene with multiple objects and a natural background of a laboratory, was considered. Figure 6.5 compares the depth maps obtained from different feature block pre-selection threshold parameters ($FB_{th}$) without using post-processing to recover the invalid results. It is seen that increasing the feature block pre-selection threshold can increase the untraceable regions been detected. This can lead to a good detection for the background, as shown in Figure 6.5(e) and (f). However, increasing of the feature block pre-selection threshold results in more regions within the object space becoming untraceable and the big untraceable region detected within object can not be properly recovered from the neighbourhood. In the current task, $FB_{th}$ =3 is chosen for a good viewing effect on the depth map, though there is some sacrifice of the neatness in the background.



(a) $FB_{th}=0$

(b) $FB_{th}=1$

(c) $FB_{th}=2$

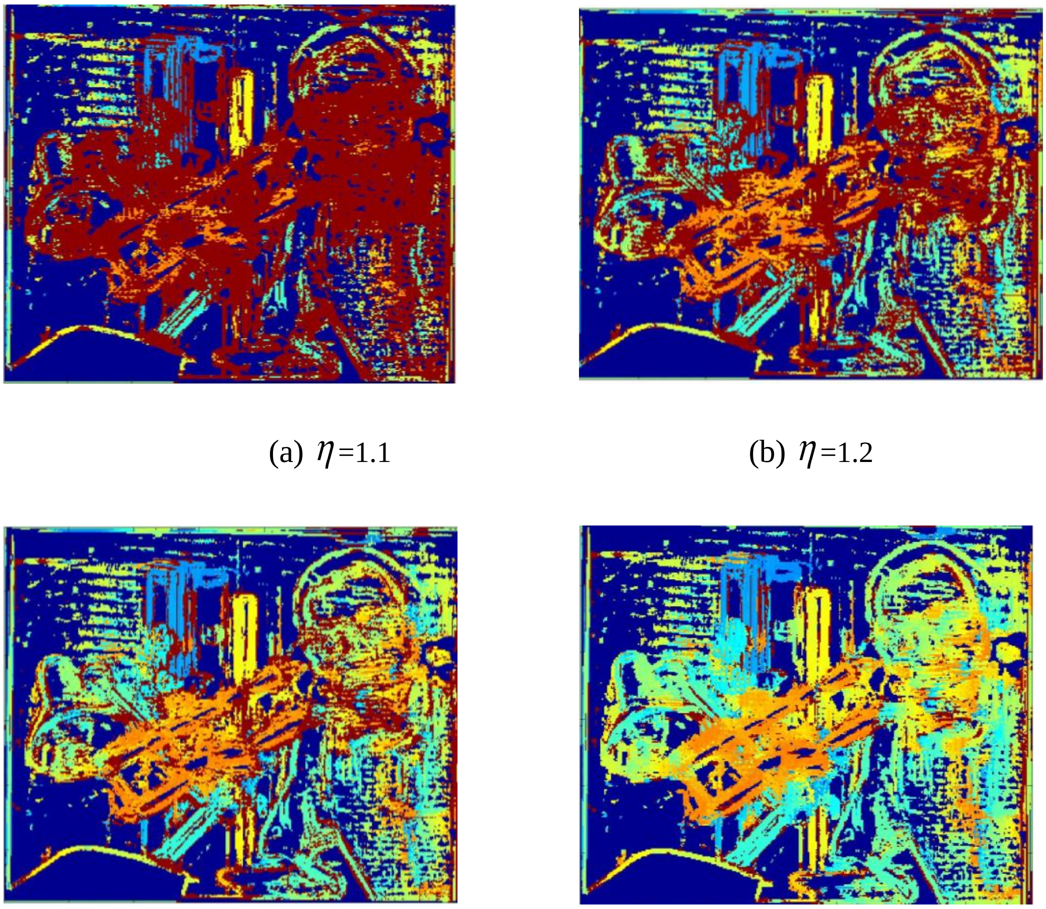
(d) $FB_{th}=3$

(e) $FB_{th}=4$

(f) $FB_{th}=5$

Figure 6.5: Comparison of the depth maps obtained from different feature block pre-selection thresholds on the captured UII data (Lab).

The effect of the consistency post-screening technique can be illustrated by referring to Figure 6.6. The figures compare the depth map obtained using different values of residue threshold parameters ($\eta$). The untrackable regions are marked in dark red in the depth map. With the increase of the residue threshold parameter ($\eta$), less positions are detected as untrackable. From the results, $\eta$=1.3 appears to be a suitable parameter in this case.



(a) $\eta$=1.1

(b) $\eta$=1.2

(c) $\eta$=1.3

(d) $\eta$=1.5

Figure 6.6: Comparison of the depth maps obtained from different residue threshold parameters on the captured UII data (Lab).

Figure 6.7 shows the final depth map obtained from Figure 6.6(c) after the invalid region has been removed (large area of untraceable region was kept intact as the background). Compared with previous result in Figure 5.26(c), the objects are clearly protrude from the screen plane and distinguished from the background.

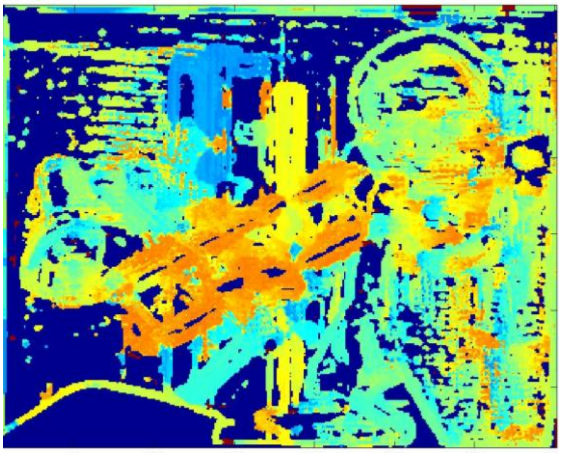


Figure 6.7: The depth map obtained from the hybrid algorithm on the captured UII data (Lab)

Figure 6.8 ~6.10 shows the final depth maps obtained from other UII data generated in the group (matchbox, horseman, and tank) respectively using the parameters listed in Table 6.2. Obvious improvement can be found in the results, especially on the images that contain regions lacking in detail. The horseman and the tank can now be easily recognized from the depth maps with the background marked out. The depth variations within objects are perceivable. For example, in figure 6.9, the horse is standing in a position with the head towards the camera, the depth increasing from the head to the tail. In figure 6.10, the tank is standing on the middle of the terrain. The picture was taken from the upper-front corner position of the tank. Details of the scenes are shown in Figure 5.26.

Table 6.2: The parameters used in obtaining Figure 6.7 ~6.10

| | Matchbox | Horseman | Tank | Lab |
|---|---|---|---|---|
| Viewpoint image numbers (N) | 12 | 8 | 8 | 12 |
| Valid viewpoint image numbers (VN) | 7 | 4 | 5 | 8 |
| Basic searching range (R) | 3 | 4 | 4 | 4 |
| Basic block Matching window size (b*w*) | 10 | 5 | 4 | 4 |
| Residue threshold ( $R_{th}$ ) parameter η | 2.0 | 2 | 1.2 | 1.3 |
| Neighbourhood shape number (NSN) | 8 | 8 | 4 | 12 |
| Neighbourhood weight factor (NWF) | 0.5 | 0.8 | 0.8 | 0.8 |
| Feature block pre-selection threshold | 4 | 6 | 4 | 3 |

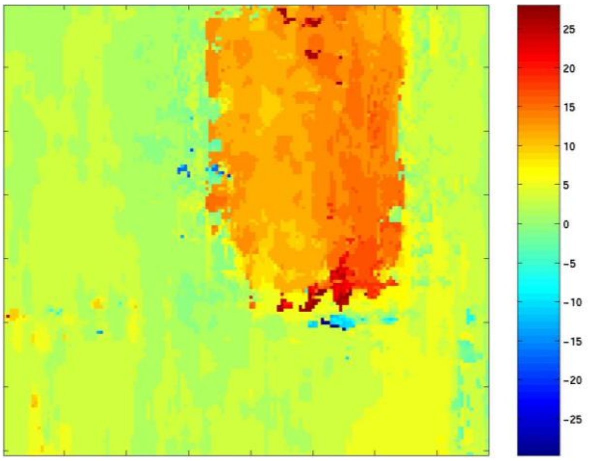


Figure 6.8: The depth map from captured UII data, matchbox.

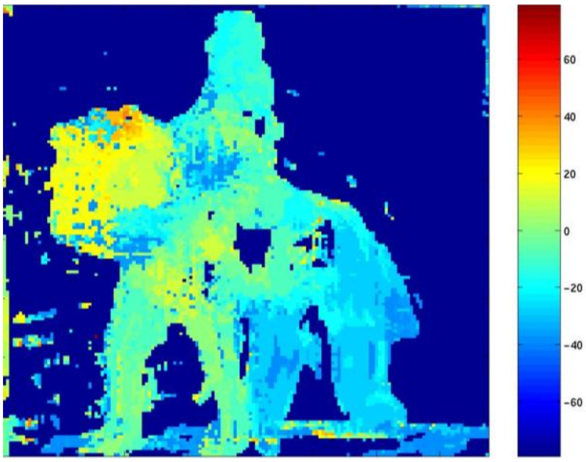


Figure 6.9.: The depth map from captured UII data, horseman.

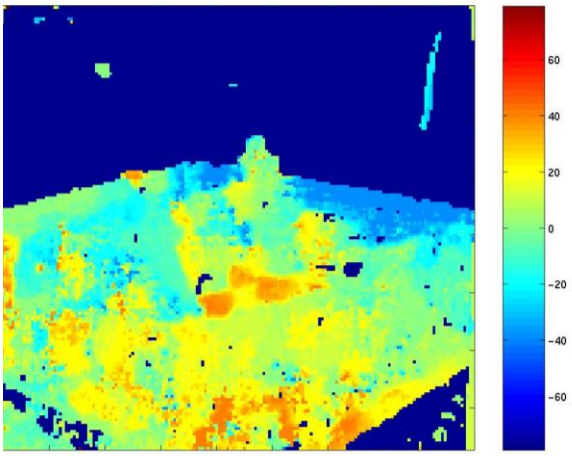


Figure 6.10: The depth map from captured UII data, tank

# 6.5 Summary

Two additional techniques--- “feature block pre-selection” and “consistency post-screening” are adopted to further improve the performance of the disparity analysis. Experiments have proved that the algorithm works effectively in detecting those positions that have difficulty in obtaining a correct match. The conclusion from the experiments has shown to be beneficial in ruling out and recovering the invalid disparity results that exist in the disparity map. Combined with the techniques used in previous chapters, an improvement can be achieved on the final depth map following the removal of the invalid results. The conclusion to be drawn from the experiments is that the overall performance of the algorithm is improved and good depth maps can be generated from both the computer generated and the captured UII data.

# Chapter 7

# Conclusions and Further Work

## 7.1 Conclusions

The present research work has approached the task of decoding the depth information embedded in a planar recording of a 3D-integral image. For simplicity, only UII data is presented and analysed but the results and approaches can be easily applied to OII data.

The 3D integral camera system and the associated image formation and recording process have been presented. The unique characteristics of the 3D-integral image data has been analysed and it has been shown that high correlation exists between the pixels at one microlens pitch distance interval. A new way of analysis 3D integral image has been established through viewpoint image extraction.

The viewpoint image extracted from the integral image is different from the traditional camera captured 2D image. An integral 3D imaging system is therefore fundamentally different from a multi-view display system, in which multiple 2D images are taken using traditional cameras placed at different positions. Similar replay results can only be achieved using a multi-view system when the object scene is distance ($\infty$) from the optical center of the recording cameras and enough image views are taken.

Viewpoint image extraction enables many well-established 2D image analysis and processing techniques to be applied to the 3D integral image with little or no adjustment.

Viewpoint image extraction explores a practical way of obtain depth information from the 3D integral data.

The main theme of the work is depth extraction from UII data. The depth equation which gives the mathematical relationship between object depth and the corresponding displacement among viewpoint images has been derived through geometric optical analysis of the integral recording. The depth extraction task has been converted to a disparity analysis task. The depth map of a 3D integral image has been estimated with an acceptable error range using a correlation-based block matching analysis on the extracted viewpoint images

To improve the ability to more accurately measure depth in 3D UII's, a number of existing methodologies have been considered. These have been modified to create improved and appropriate UII depth extraction algorithms.

Initially, a multi-baseline technique has been adapted to overcome the ambiguity that exists in disparity analysis due to the similar patterns within the object scene. In the multi-baseline technique, the matching judgment is carried out using an accumulated score function from all the viewpoint image pairs with different baselines. This enables a robust and precise measurement to be made by taking advantage of information redundancy contained in multiple images of the same scene. Modifications to an original multi-stereo algorithm have been carried out to accommodate the fact that the viewpoint image is generated in a different way from a traditional 2D image and a new depth equation has been formulated to account for the difference in the definition of the baseline between stereo and integral images. The effectiveness of the modified multi-baseline algorithm has been proved both mathematically and experimentally. Accurate depth measurement has achieved on the testing UII data and depth maps and object scene reconstruction have been carried out on both computer generated and camera captured UII data.

A neighbourhood constraint and relaxation technique has been applied to further improve the disparity analysis performance. This was instituted in an effort to obtain a good trade-off in choosing a suitable window size for matching. The algorithm uses a complicated evaluation criterion in considering the spatial constraint. The neighbourhood of the block is involved in the evaluation with suitable weighting factors and flexible disparity variance allowance. In general, a higher weighting factor is assigned to a block which has the great possibility of being in the same object as the feature block. Within the neighbourhood constraint and relaxation algorithm, a two-stage voting scheme with a multi-candidate pre-screening strategy is introduced to improve the computation efficiency. The multi-candidate pre-screening used in the first stage largely reduces the number of computations in the second stage of decision making. Experiments showed an obvious improvement in the depth map generated from the algorithm.

The untraceable homogenous region and untrackable region around object borders are two very difficult situations for disparity analysis to deal with. To enable correct depth estimation within these regions “feature block pre-selection” and “consistency post-screening” techniques have been introduced. The untraceable region and untrackable region are detected either before or after the matching analysis. The depths of these positions can be roughly recovered by considering the neighbouring positions using spatial constraints. The large homogeneous region in the image is recognized as background. The final depth maps obtained from the available UII data are visually satisfactory and represent a marked improvement when compared to previous results.

In a conclusion, a new way of analyzing 3D integral imaging system through viewpoint image extraction has been established. A new depth equation, which describes the mathematical relationship between object depth and corresponding viewpoint images displacement has been developed and depth maps have been generated from realistic II’s through disparity analysis. The developed hybrid disparity analysis algorithm with the feature block pre-selection and consistency post-screen technique is not confined to analysis the disparity on viewpoint images. The principle can be used in analyzing the disparity from a sequence of 2D images for obtaining spatial or motion information of objects in a scene.

## 7.2 Further work

1) A number of techniques might further improve the performance of the disparity analysis. For examples:

I) Within the neighborhood constraint and relaxation algorithm described in the thesis, a fine weight scheme could be considered for identifying color/texture similarity and lead to an improved estimate of the confidence gained from the disparity matching result.

II) The hybrid disparity analysis algorithm involves using several techniques with different parameters. These parameters are currently manually operated according to the results achieved on the final depth map. In addition, the performance of the depth extraction is subjectively evaluated. It is proposed that automatic operating of the parameters could be achieved by setting up a cost function as an objective evaluation function. With the objective evaluation standard, either simulated annealing or genetic algorithms maybe adapted to adjust the parameters to achieve improved results for the output depth map.

III) Considering the basic matching window size used in matching, an adaptive window could be used by evaluating the local variation of the intensity and disparity (Kanade and Okutomi1990, Okutomi and Kanade 1991).

IV) Current work has concentrated on the gray-level coded UII data, where the matching is carried out by comparing the intensity variations between matching windows. By considering the colour information in the matching process, better results supposed to be achievable.

Using a RGB color model, the matching evaluation equation can be written as:

$$SSD(D) = \sum_{x,y \in w} \sum_{k \in (r,g,b)} [\hat{I}^k{}_1(x,y) - \hat{I}^k{}_2(x+D,y)]^2$$

2) Current work on depth extraction has concentrated on UII data involving only 1D space in the disparity analysis. The same approach can be used to extract the depth map from OII data by considering the disparity that exists in 2D space.

3) A segmentation algorithm to divide the 3D image into different 3D object planes according to the different object depth should be developed to enable content-based image coding and content-based interactive manipulation for Virtual Studios.

A proposed 3D II decomposing scheme is given in figure 8.1. Along the red arrow, the 3D II which contains multiple objects can be decomposed into several 3D II plane (3D II P1 & 3D II P2), each 3D II plane only contains one object in a particular depth. Therefore, each object can be manipulated separately according to the requirement. For example, different objects from several 3D II either computer generated or photographically captured can be combined or removed flexibly in a virtual studio.

Along the green arrow, each 3D II plane can be further decomposed into sub-viewpoint images (SVI). Each sub-viewpoint image is a 2D parallel recording of a particular object. With the depth information obtained for each object plane, each 3D II plane could be coded as one sub-viewpoint image plus the corresponding depth coding for the plane. Hence, an efficient compression scheme can be achieved for the content-based coding scheme. In the proposed content-based coding scheme, each 3D II is now represented by a number of 2D images with different objects, this will enable easy content-based image retrieval which is essential in searching and retrieval from huge images files.

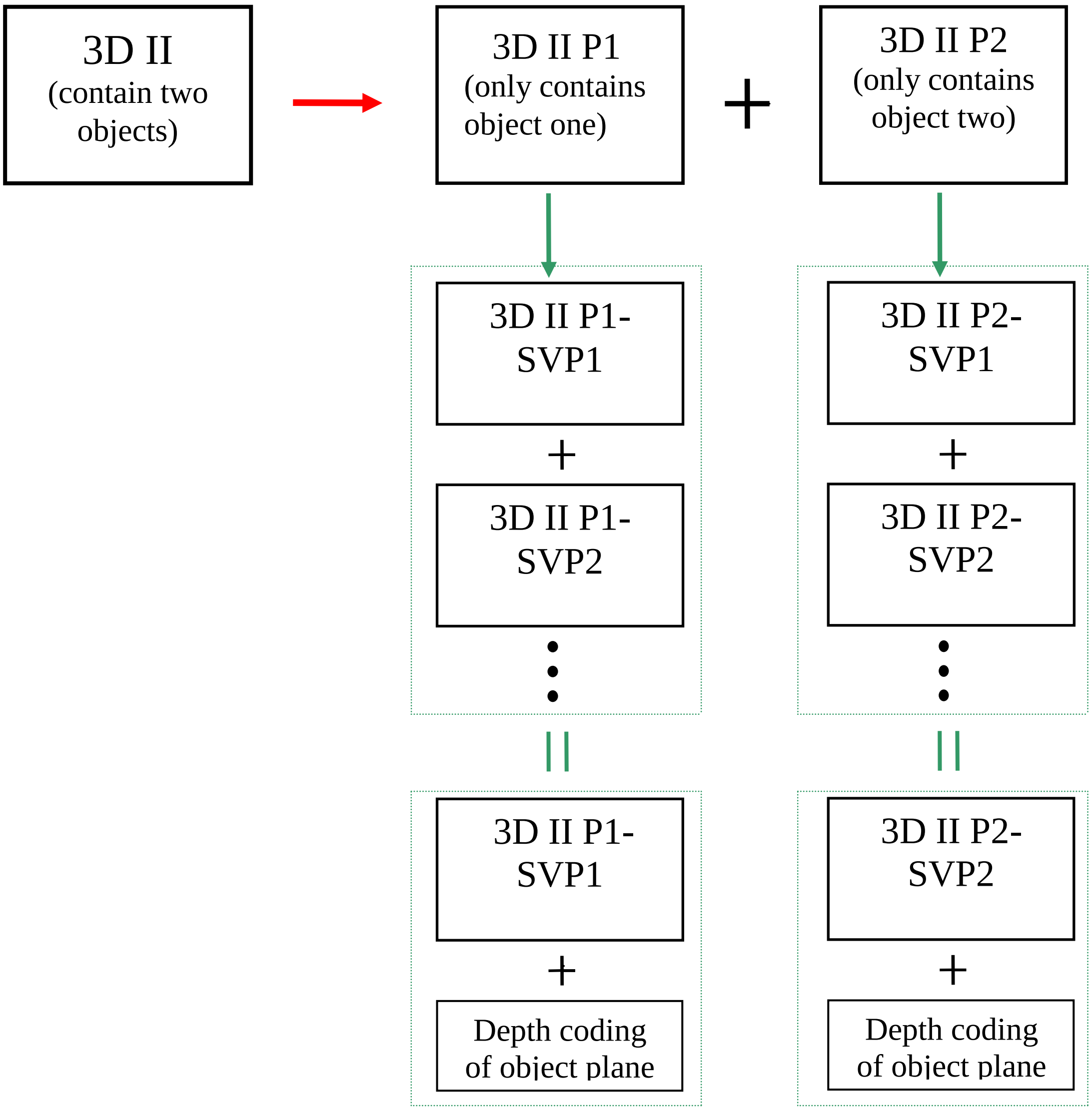


Figure 7.1: The proposed 3D II decomposing model

(For illustration, assume only two objects exist in the scene)

# References

[3DCGI] http://www.3dcgi.com/coltech/displays/displays.html

[4D-Vision] 4D-Vision GmbH, http://www.4d-vision.de/

[Actualdepth] Deep Video Imaging Ltd, http://www.deepvideo.com/index3.html

[Actuality Systems] Actuality Systems, Inc., http://www.actuality-systems.com/

[Aggarwal 1988] Aggarwal J and Nandhakumar N, "On the Computation of Motion from sequences of images-A review", *Proceedings of the IEEE*, vol. 76(8), pp917-935, 1988.

[Amateur Holography] http://members.aol.com/gakall/holopg.html

[Appel 1968] Appel A, "Some techniques for shading Machine renderings of solids", *Spring Joint Compter Conference*, pp37-45, 1968.

[Arai 1998] Arai J, Okano F, Hoshino H and Yuyama I, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images", **App. Optics**, vol. 37(11), pp2034-2045, 1998.

[Barnard 1989] Barnard S, "Stochastic stereo matching over scale", **Int. J. Comput. Vision**, pp.17-32, 1989.

[Benton1980] Benton, S., *US Patent* US4431265, "Apparatus for viewing stereoscopic images", 1980

[Benton1980] Benton, Stephen A, "Holographic displays: 1975-1980", **Opt Eng**, vol. 19(5), pp686-690, 1980.

[Bertero 1998] Bertero, M., Boccacci,P., **Introduction to inverse problems in imaging**, Institute of Physics Publishing , Bristol and Philadephia, 1998.

[Bierling 1988] Bierling M, “Displacement estimation by hierarchical blockmatching”, *Proc. SPIE Visual Communications and Image Processing*, vol. 1001, pp942-951, 1988.

[Bourke1999] Bourke P, “Autostereoscopic lenticular images”, http://astronomy.swin.edu.au/~pbourke/stereographics/lenticular/

[Burkhardt 1969] C.B.Burkhardt and E.T.Doherty, “Beaded plate recording of integral photographs,” **App. Optics** 8(11), pp. 2329-2331, 1969.

[Cartwright 2000] Cartwright P, **Realisation of computer generated integral three dimensional images**, De Montfort University, Thesis, 2000.

[Chang 2000] Chang P and Wu M, “A wavelet multiresolution compression technique for 3D stereoscopic image sequence based on mixed-resolution psychophysical experiments”, **Signal Processing: Image Communication**, vol.15(9), p705-727, July 2000.

[Chen 1990] Chen J and Medioni G, “Parallel multiscale stereo matching using adaptive smoothing”, *ECCCV90*, 1990.

[Chen 1996] Chen Y, Lin Y and Kung S: “A feature tracking algorithm using neighborhood relaxation with multi-candidate pre-screening”, *Proc. of International Conference on Image Processing*, vol. II, September, 1996.

[Chen 1998] Chen Y, **True Motion Estimation-Theory, Application, and Implementation,** Princeton University, Thesis, 1998.

[D4D] Dresden 3D GmbH, http://www.dresden3d.com/

[Dallaert 2002] Dellaert F, Seitz S, Thorpe C and Thrun S, “Structure from motion without correspondence”, *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2000.

[Davies 1988] Davies N, “Three dimensional imaging systems: A new development”, **Appl. Optics**, vol. 27, pp. 4520-4528, 1988

[Davies 1992] Davies N and McCormick M "Holoscopic Imaging with True 3-D Content in Full Natural Colour", **Journal of Photographic Science**, vol. 40, pp.46-49, 1992.

[Davies 1994] Davis N, McCormick M, and Brewin M, "Design and analysis of an image transfer system using microlens arrays", **Optical Engineering**, vol. 33, no. 11, pp. 3624-3633, 1994.

[DDD] Dynamic Digital Depth company, http://www.ddd.com/

[Dhond 1989] Dhond U and Aggarwal J, "Structure from stereo - a review", **IEEE Transactions on Systems, Man and Cybernetics**, vol.19(6), 1989.

[Dodgson 1997] Dodgson, "Autostereo displays: 3D without glasses", *EID'97*, Surrey.

[DTI] Dimension Technologies, Inc, http://www.dti3d.com/

[Dufaux 1995] Dufaux F and Moscheni F, "Motion Estimation Techniques for Digital TV: A Review and a new contribution", **Proceedings of the IEEE**, vol. 83, no. 6, pp. 858-876, 1995.

[Forman 1999] Forman M, **Compression of Integral Three-dimensional Television Pictures**, De Montfort Univesity, Thesis, 1999.

[Forstner 1986] Forstner W and Pertl A, "Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements", **Pattern Recognition in Practice II,** pp57-72, 1986.

[Gabor1948] Gabor, D., "A new microscope principle", **Nature**, No. 161, pp. 777-779, 1948

[Gabor1949] Gabor, D., "Microscopy by reconstructed wavefronts", **Pro. Phys. Soc**. A194, pp454-487, 1949.

[Genex] Genex$^{TM}$**,** http://www.genextech.com/

[Hannah 1989] Hannah M, “A system for digital stereo image matching”, **Photogram. Eng. Remote Sensing,** vol. 55, no. 12, pp.1765-1770, 1989.

[Hariharan2002] Hariharan P, **Basics of Holography**, Cambridge university press, 2002

[Harman 1996] Harman P, “ Autostereoscopic display system”, Proceeding of SPIE, vol. 2653, pp56-64

[Harou 1992] Harou, I. and Minoru, Y., “50-inch Autostereoscopic Full Colour 3D TV Display System”, *SPIE 1669,* 176-179, 1992

[Huang 1994] Huang T and Netravali A, “Motion and structure from feature correspondences: A review”, *Proceedings of the IEEE*, 82(2), 1994.

[Ives 1931] Ives H E: “Optical Properties of a Lippmann Lenticulated Sheet”, *J. Opt. Soc. Amer*., **vol.21**, pp171-176,1931.

[Ives1903] Ives, U.S. Patent No. 725,567, (1903).

[Javidi 2001] Javidi B, Min S and Lee B, “Enhanced 3D color integral imaging using multiple display devices”, *Proceeding of IEEE Lasers and Electro-Optics Society Annual Meeting,* p491-492, 2001

[Jebara 1999] Jebara T, Azarbayejani A and Pentland A, "3D And Stereoscopic Visual Communication", IEEE Signal Processing, vol. 16. No. 3, 1999.

[Kanade 1994] Kanade T and Okutomi M, “A stereo matching algorithm with an adaptive window: Theory and experiments*”, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920-- 932, September 1994.

[Kasper 1987] Kasper J and Feller S, **The Complete Hologram Boo**k, Prentice-Hall, 1987.

[Kishigami 2001] Kishigami R, Takahashi H, Shimizu E: “Real-time color three-dimensional display system using holographic optical elements”, *Proceedings of SPIE* vol.4296, pp102-107, 2001.

[Kratomi1972] Kratomi, S., US patent US3737567, “Stereoscopic Apparatus Having Liquid Crystal Filter Viewer”, 1972.

[Langhans 2002] Langhans, K. etc, “FELIX 3D Display: An Interactive Tool for volumetric Imaging”, Proceedings of SPIE, vol. 4660, Stereoscopic Displays and Virtual Reality Systems IX”

[Lee 2002] Lee B, Min S and Javidi B, “Theoretical analysis for three-dimensional integral imaging systems with double devices”, **Applied Optics,** vol. 41(23), pp4856-4865, 2002.

[Leith1963] Leith E and Upatnieks J, “Wavefront reconstruction with continuour-tone objects*”,* **J. Opt. Soc. Amer**., vol. 53(12), pp. 1377-1381,1963.

[Lewis1971] Lewis, Jordan D, Verber, Carl M and McGhee A, “True three-dimensional display”, **IEEE Trans Electron Devices**, vol. 18, pp724-732.

[Li 2002] H.Li, R.Sannino and J. kari: “A multi-stage block matching motion estimation for super-resolution video reconstruction”, *Proceedings of SPIE 2002*, January, San Jose, USA.

[Lippmann 1908] Lippmann G: “La Photographie integrale”, **comtes Rendus**, *Academie des Sciences*, vol.146, pp.446-451,1908.

[Lucas 1981] Lucas B, Kanade T, “An iterative image registration technique with an application to stereo vision”, *Image understanding workshop,* pp121-130, 1981.

[Manolache 1999] Manolache S, Aggoun A, McCormick M and Davies N, “A mathematical model of a 3D-lenticular integral recording system”, *Proceedings of IEEE Vision, Modeling and Visualization Conference,* Erlangen, pp 51-58, 1999.

[Manolache 2001] Manolache S, McCormick M and Kung SY: “Hirearchical adaptive regularization method for depth extraction from planar recording of 3D-integral images”, *IEEE Proceedings of ICASSP*, vol. 3, pp1433-1436, 2001.

[Manolache 2002] Manolache S, Kung SY, McCormick M and Aggoun A: “3D-object space reconstruction from planar recorded data of 3D-integral images”, **J. VLSI Signal Processing Systems,** Kluwer Academic Publishers, 2002.

[Matthies 1989]Matthies L, Szeliski R and Kanade T, “Kalman filter-based algorithms for estimating depth from image sequences”, **Int. J. Comput. Vision**, vol. 3, pp.209-236,1989.

[McAllister 1993] McAllister D, **Stereo computer graphics and other true 3D technologies**, Princeton university press, 1993.

[McCormick 1992] McCormick M and Davies N, “3-D worlds”, **Physics World**, pp.42-46, June 1992.

[McCormick 1992] McCormick M, Davies N and Chovanietz E., “Restricted parallax images for 3D T.V.”, *Proc. IEE, Conf. Pub.* No. 172, London, 1992

[McCormick 1992] McCormick M, Davies N and Chovanietz E., “Towards Continuous Parallax3-Images”, *IEE Colloq. ‘Stereoscopic Television’*, no. 173, 3/1-3/4, Nov.1992

[McCormick 1994] McCormicK M, “Examination of the requirements for autostereoscopic, full parallax, 3D TV”, *Conf. Pub. IEE*, vol. 397, pp. 477-482, 1994

[McCormick 1995] McCormick M and Davies N, “Full natural colour 3D optical models by integral imaging”, *Proc. IEE, 4th Int. Conf. On Holographic Systems, Components and Applications*, No. 379, pp. 237-242, Switzerland, 13th-15th Sept. 1995

[McCormick 1995] McCormick M, “Integral 3D imaging for Broadcast”, *Proc. of Second International Display Workshops*, vol. 3, pp. 77-80, 1995.

[Min 2001] Min S, Jung S, Park J, Lee B, “Three-dimensional display system based on computer generated integral photography”, *Proceedings of SPIE*, vol.4297, pp187-195, 2001.

[Mohr 1996] Mohr R and Tiggs B, “Projective geometry for image analysis”, Technical report, *International Society for Photogrammetry and Remote Sensing, Vienna Congress*, July,1996.

[Mori 1973] Mori K, Kidode M and Asada H, “An iterative prediction and correction method for automatic stereocomparison”, **Computer graphics and image processing**, vol.2, pp393-401, 1973.

[Motoki 1995] Motoki T, Isono H and Yuyama I, “Present status of three-dimensional television research”, **Proc. IEEE83**, pp. 1009-1021, 1995.

[Naeumra2001] Naemura T, Yoshida T and Harashima H, “3-D computer graphics based on integral photography”, **Optics express**, vol.8,No.2, pp255-262,2001.

[Okano 1998] Okano F, Hoshino H, Arai J and Yuyama I, “Real-time pickup method for a three-dimensional image based on integral photography”, **Applied optics**, vol.36, No. 7, pp1598-1603.

[Okoshi1976] Okoshi T, **Three dimensional imaging Techniques**, Academic Press, London, 1976.

[Okutomi 1993] Okutomi M and Kanade T, “A Multiple-Baseline Stereo”, **IEEE Transactions on Pattern Analysis and Machine Intelligence,** vol. 15(4), 1993, pp. 353-363.

[Orchard 1993] Orchard M, “Predictive Motion-Field Segmentation for Image Sequence Coding”, **IEEE Transactions on circuits and systems for Video Technology**, vol. 3(1), pp54-70, Feb. 1993.

[Outwater1999] Outwater C and hamersveld V, “Practical Holography”, Dimensional Arts Inc. 1995-99, http://www.holo.com/holo/book/book1.html.

[Philips] Philips,http://www.research.philips.com/

[Poelman 1993] Poelman C and Kanade T, “A paraperspective factorization method for shape and motion recovery”, *Tech. report CMU-CS-93-219*, Computer Science Department, Carnegie Mellon University, December, 1993.

[Povray 1999] POV-Ray Version 3.00, http://www.povray.org/ persistence of vision ray tracer.

[Puri 1997] Puri A, Kollarits R and Haskell B. “Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4”, **Signal Processing: Image Communication**, vol.10.1-3, p201-234, July 1997.

[Schmidt 2002] Schmidt A and GrasnickA.: “ Multi-viewpoint Autostereoscopic Displays from 4D-Vision”, *Proceedings of SPIE* ,vol. 4660, p212-221, 2002.

[Siegel 1995] Siegel M, Grinberg V, Jordan A, McVeigh J,Podnar G, Safier S and Sriram S, “ Software for 3D-TV and 3D-Stereoscopic Computer Workstations”, *Proc. of the International Workshop on Stereoscopic and Three Dimensional Imaging*, pp. 251 - 260, 1995.

[Srinivasan 2000] Srinivasan S, “ Extracting structure form optical flow using the fast error search technique”, *International journal of computer vision*, vol. 37(3), pp203-230, 2000.

[Stereographics] Stereographics, http://www.stereographics.com/

[Stevens2001] Stevens R, Davies N. and Milnethorpe G, “ Lens arrays and optical system for orthoscopic three-dimensional imaging”, **The Imaging Science Journal**, vol. 49, pp151-164.

[Taub2002] Taub A. “Will 3-D ever catch on?”, The New York Times.

[Tetsutani1994] Tetsutani, N., Omura, K., Kishino, F., “Wide-screen autostereoscopic display system employing head-position tracking”, **Opt. Eng.,** vol. 33, no. 11, pp. 3690-3697, November 1994.

[Trucco 1998] Trucco E**, Introductory techniques for 3-D computer vision**, Prentice Hall, 1998.

[Valyus1966] Valyus N, **Stereoscopy**, Focal Press, London 1966.

[Wang 2002] Wang, W, Yao Q and Yu L: "Fast algorithm of arbitrary fractional-pixel accuracy motion estimation", *Proceedings of SPIE,* 2002.

[Wu 2002] Wu C, Aggoun A, McCormick M and Kung SY: "Depth extraction from unidirectinal integral image using a modified multi-baseline technique", *Proceedings of SPIE*, vol.4660, pp135-145, Jan, 2002.

[Yano 2002] Yano S, Ide S. Mitsuhashi T and Thwaites H, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images", **Displays**, vol.23 (4) PP191-201, 2002.

[Zaharia 2001] Zaharia R, **Adaptive Compression Algorithm for Full Colour Three Dimensional Integral Images**, Thesis, De Montfort University, 2001.

# Appendix: The Test Integral Image data

The intensity distributions presented in this appendix are those which have been used in generated depth map described in this thesis.  They have been reproduced at the correct scale for three dimensional display use the lenticular sheets provided.  The appropriate lenticular sheet for each intensity distribution is given in its caption.

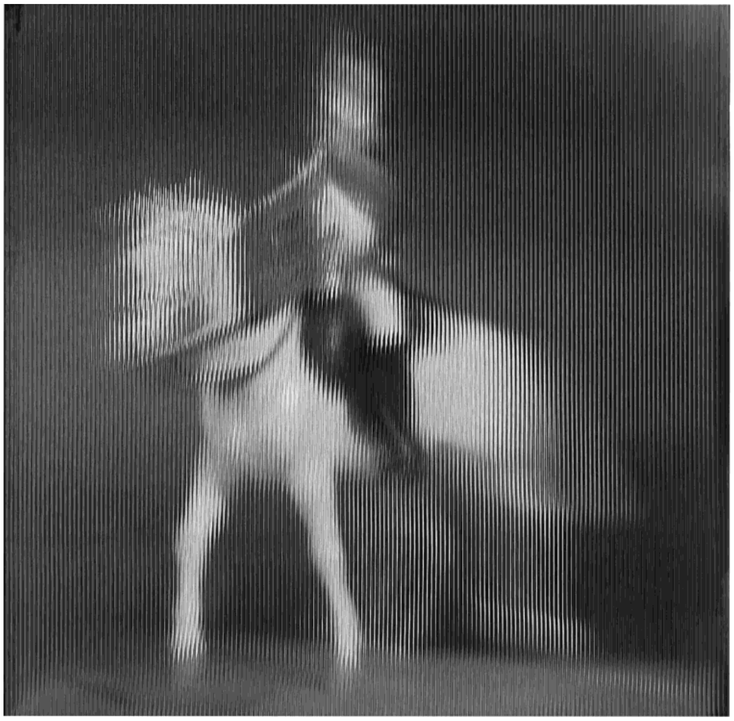


Image A.1:  Horseman (600*um*)

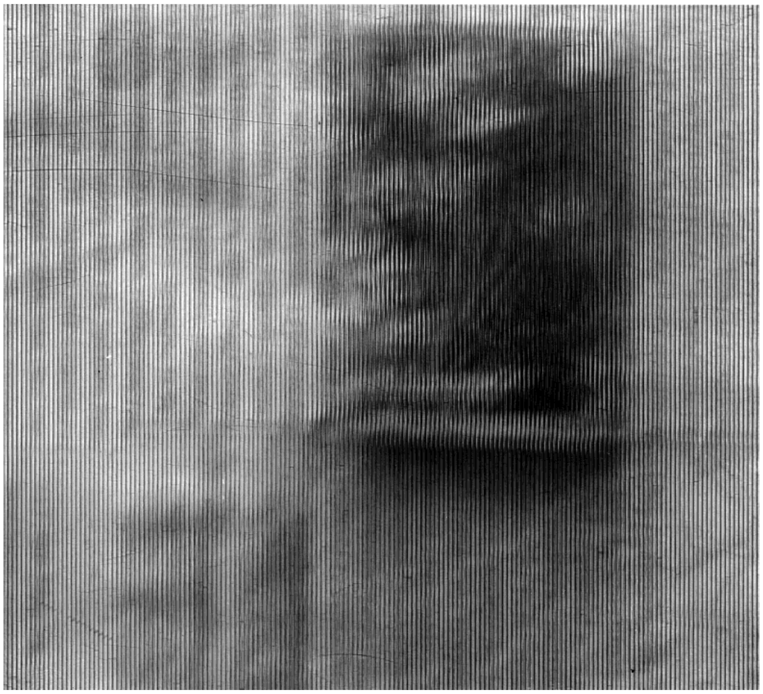


Image A.2:  Matchbox (600*um*)

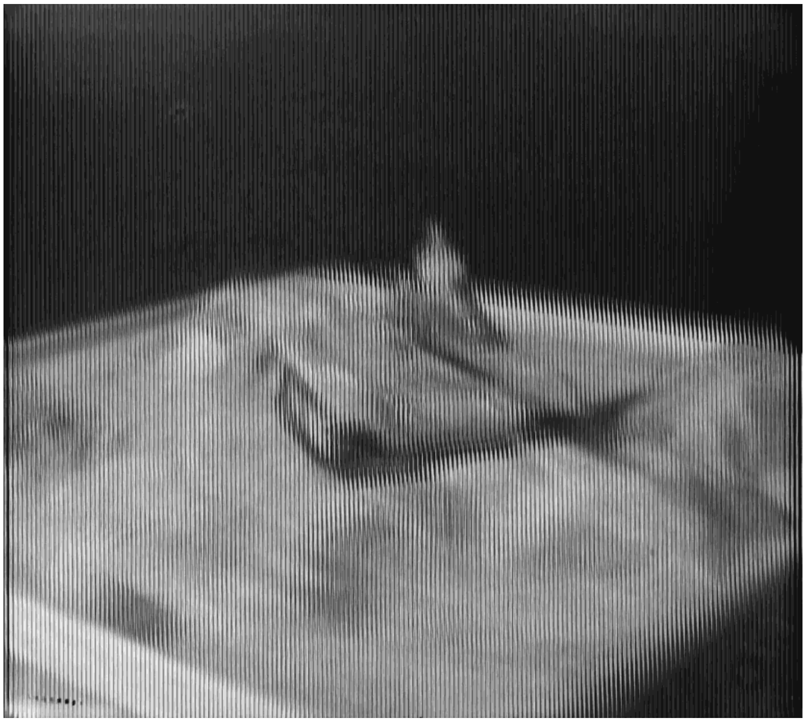


Image A.3: Tank (600*um*)

Image A.4: Lab (600*um*)

Image A.4: Lab (600*um*)

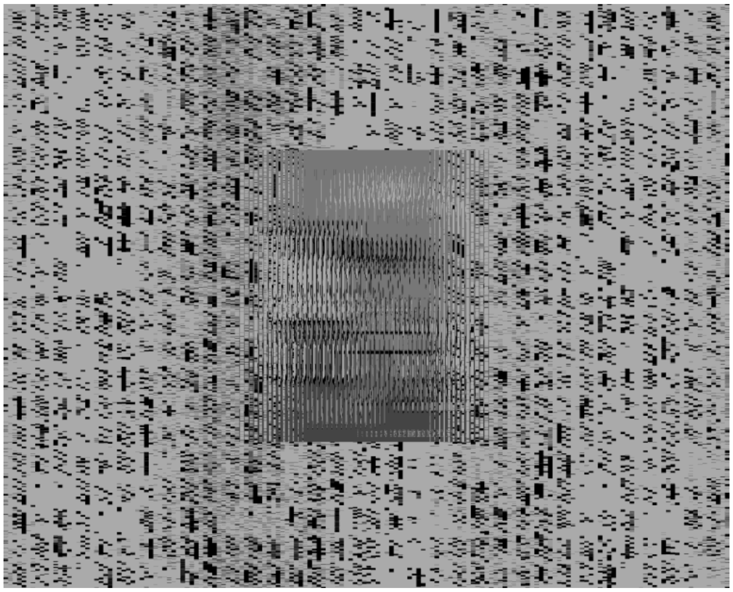


Image A.5:  cg_box (600*um*)

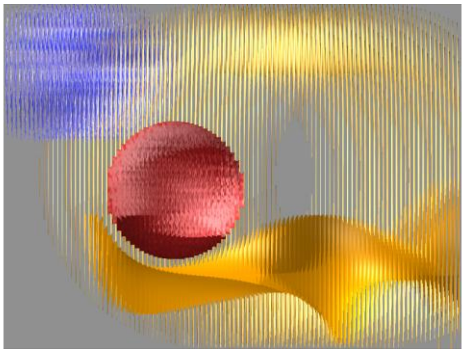


Image A.6:  cg_balls (600*um*)