Holoscopic 3D Image Depth Estimation and Segmentation Techniques

by

Eman F. M. Alazawi

A thesis submitted for the degree of

Doctor of Philosophy

in

Department of Electronic & Computer Engineering

Collage of Engineering, Design and Physical Sciences

Brunel University London

January 2015

ABSTRACT

Today's 3D imaging techniques offer significant benefits over conventional 2D imaging techniques. The presence of natural depth information in the scene affords the observer an overall improved sense of reality and naturalness. A variety of systems attempting to reach this goal have been designed by many independent research groups, such as stereoscopic and auto-stereoscopic systems. Though the images displayed by such systems tend to cause eye strain, fatigue and headaches after prolonged viewing as users are required to focus on the screen plane/accommodation to converge their eyes to a point in space in a different plane/convergence.

Holoscopy is a 3D technology that targets overcoming the above limitations of current 3D technology and was recently developed at Brunel University. This work is part W4.1 of the 3D VIVANT project that is funded by the EU under the ICT program and coordinated by Dr. Aman Aggoun at Brunel University, West London, UK. The objective of the work described in this thesis is to develop estimation and segmentation techniques that are capable of estimating precise 3D depth, and are applicable for holoscopic 3D imaging system. Particular emphasis is given to the task of automatic techniques i.e. favours algorithms with broad generalisation abilities, as no constraints are placed on the setting. Algorithms that provide invariance to most appearance based variation of objects in the scene (e.g. viewpoint changes, deformable objects, presence of noise and changes in lighting). Moreover, have the ability to estimate depth information from both types of holoscopic 3D images i.e. Unidirectional and Omni-directional which gives horizontal parallax and full parallax (vertical and horizontal), respectively.

The main aim of this research is to develop 3D depth estimation and 3D image segmentation techniques with great precision. In particular, emphasis on automation of thresholding techniques and cues identifications for development of robust algorithms. A method for depth-through-disparity feature analysis has been built based on the existing correlation between the pixels at a one micro-lens pitch which has been exploited to extract the viewpoint images (VPIs). The corresponding displacement among the VPIs has been exploited to estimate the depth information map via setting and extracting reliable sets of local features.

Feature-based-point and feature-based-edge are two novel automatic thresholding techniques for detecting and extracting features that have been used in this approach. These techniques offer a solution to the problem of setting and extracting reliable features automatically to improve the performance of the depth estimation related to the generalizations, speed and quality.

Due to the resolution limitation of the extracted VPIs, obtaining an accurate 3D depth map is challenging. Therefore, sub-pixel shift and integration is a novel interpolation technique that has been used in this approach to generate super-resolution VPIs. By shift and integration of a set of up-sampled low resolution VPIs, the new information contained in each viewpoint is exploited to obtain a super resolution VPI. This produces a high resolution perspective VPI with wide Field Of View (FOV). This means that the holoscopic 3D image system can be converted into a multi-view 3D image pixel format. Both depth accuracy and a fast execution time have been achieved that improved the 3D depth map.

For a 3D object to be recognized the related foreground regions and depth information map needs to be identified. Two novel unsupervised segmentation methods that generate interactive depth maps from single viewpoint segmentation were developed. Both techniques offer new improvements over the existing methods due to their simple use and being fully automatic; therefore, producing the 3D depth interactive map without human interaction.

The final contribution is a performance evaluation, to provide an equitable measurement for the extent of the success of the proposed techniques for foreground object segmentation, 3D depth interactive map creation and the generation of 2D super-resolution viewpoint techniques. The no-reference image quality assessment metrics and their correlation with the human perception of quality are used with the help of human participants in a subjective manner.

ACKNOLEGMENT

I would like to express my sincere appreciation and thank to all those who made this thesis possible. This work would not have been possible without help, support and motivation of my supervisors, Dr Maysam Abbod and Prof John Cosmas.

I am heartily thankful to Prof A. Aggoun for giving me the opportunity to work with him during his 3D VIVANT research project and his encouragement, guidance, and support from the beginning till the end made my work much easier. The support I have received from 3D VIVANT project team has been essential for the good development of this work. I thoroughly enjoyed working in the field of Holoscopic 3D Imaging system and it has been a very fruitful experience during my research and I can anticipate a positive impact on my research in future.

Working at 3D VIVANT has been tremendously pleasant thanks to all members. In particular, I would like to warmly thank M. R. Swash for the numerous comments on my work and feedback he gave me. I would also like to thank O. Abdul Fatah for the endless discussions about the area and the work under taken.

I am indebted to show my gratitude to Carlos, who has supported me throughout capturing holoscopic 3D images during my research and Bie Li for his useful and professional comments.

Last but not least, my sincere appreciation goes to all of those who supported me during my studies.

STATEMENT OF ORIGINALITY

The work contained within this thesis is purely that of the author unless otherwise stated. Except what is acknowledged, none of the work presented here has been published or distributed by anyone other than the author.

Signature: ....

Eman F. M. Alazawi

January, 2015, London

TABLE OF CONTENTS

Table of Figures

Table of Tables

List of Abbreviations

1D	One-decantation
2D	Two-dimensional Image
2DVPI	Two-dimension Viewpoint Image
3D	Three-dimensional Image
3DIM	Three-dimension Interactive Map
3DTV	Three-dimensional Television
AFE	Auto-Feature-Edge
AFP	Auto Feature-Point
ANOVA	ANalysis Of VAriance
ASW	Adaptive Support Weight
AWF	Adaptive Weighting Factor
BFEP	Boundary Feature-Edge Pixel
CCD	Charge Coupled Device
CHI	Computational Holoscopic Image
CMOS	Complimentary Metal-Oxide Semiconductor
DLP	Digital Light Processing
DOF	Depth Of Field
DSCQS	Double Stimulus Continuous Quality Scale
EIs	Elemental Image/Images
FEs	Feature-Edges/Edge
FMS	Feature-Match Selection
FOV	Field Of View
fps	frames per second
FWQM	Full-Width-At-Quarter-Maximum
H3DI	Holoscopic 3D Imaging/Image
HMBA	Hybrid Multi-Baseline Algorithm
HRVPI	High Resolution Viewpoint Image
HVS	Human Visual System
IOU	Intersection-Over-Union
IP	Integral Photography

LCD	Liquid Crystal Display
LHA	Local Histogram Analysis
LRVPI	Low Resolution Viewpoint Image
MBT	Morphology-Based-Threshold
MOS	Mean Opinion Score
MQLH	Multi-Quantized Local Histogram
MRE	Mean Relative Error
MSE	Mean Square Error
MSSIM	Mean Structural SIMilarity
OH3DI	Omnidirectional Holoscopic 3D Imaging/Image
OH3DEIs	Omnidirectional Holoscopic 3D Elemental Images
OTT	Optimal Threshold Technique
PSNR	Peak Signal-to-Noise Ratio
PWF	Pixel Weight Factor
QoE	Quality of Experience
SAD	Sum of Absolute Difference
SE	Structure Element
SIFT	Scale-Invariant Feature Transform
SLMs	Spatial Light Modulators
SR	Super Resolution
SRG	Seed Region Growing
SSD	Sum of Squared Differences
SSSD	Sum of SSD
SURF	Speeded Up Robust Features
UH3DI	Unidirectional Holoscopic 3D Imaging/Image
UH3DEIs	Unidirectional Holoscopic 3D Elemental Images
VPI	Viewpoint Image

CHAPTER 1

Introduction

1.1 Preface

When analysis needs to be performed on real three-dimensional objects and their environments, a 3D study needs to be developed; and, 3D object reconstruction is a way to achieve this. Todays, 3D imaging is a new type of visual media that has greatly expanded what the viewer can see over the traditional 2D imaging technology, where, in 2D media, there are four cues missing [1]:

- Stereo parallax—seeing a different image with each eye.
- Movement parallax—seeing different images when we move our heads.
- Accommodation—the eyes' lenses focus on the object of interest.
- Convergence—both eyes converge on the object of interest.

With the development of different display systems and the requirement to delight the user with a realistic 3D world, the representation and rendering of realistic 3D impressions is a research area with excellent prospects and great challenges. 3D imaging technology has become an attractive technique that offers natural products of 3D depth and motion parallax that humans can perceive naturally as being true to the real world [2]. There is growing evidence that 3D imaging techniques will have the potential to establish a future mass-market in the fields of 3DTV (i.e. 3D display and video communication systems), 3D cinema, medical imaging, military design, robotic vision, biometrics, detection and tracking of people and in video games development through appropriate integration of real and synthetic images [3] [4] [5].

Today most 3D technologies are competing to provide a new form of 3D video objects that have natural colour, full parallax and without the need to wear any specific glasses. The first type of 3D system created the illusion of depth by looking simultaneously at different images taken from a slightly translated viewpoint of the same scene through each eye. It was invented by Sir Charles Wheatstone in June 1838 and he proposed that it should be called a "Stereoscopy", to indicate its property of representing solid figures [6]. This 3D creation technique requires special 3D glasses that filter the corresponding views for the left and right eyes of each viewer [7] [8].

Despite the disadvantages of this technique, i.e. unnatural 3D depth, single fixed perspective and eye stress, it is still used for creating depth effects in movies, computer games and product presentations, etc. In this technique, the filtering glasses are an essential requirement that enables viewers to perceive the left eye image and the right eye image with the left eye and the right eye, respectively [9]. The human brain then processes the two images and leads the viewer to believe he or she is viewing a real 3D space.

Developments in digital technology have resulted in a demand for true 3D viewing, and research groups have developed more advanced 3D display systems referred to as "auto-stereoscopic". These "no glasses" systems provide true 3D images during work procedures; therefore, they are more comfortable for the viewer as they do not require the use of special glasses [1] [10]. Holography [11], volumetric displays [12], multi-view [13] and integral imaging [14] are some types of auto-stereoscopic technology. In this technique, a combination of stereo parallax and movement parallax effects are exploited to give 3D without glasses in multi-view and head-tracking auto-stereoscopic displays [15]. All the auto-stereoscopic display systems utilize optical components to enable different images to be visible in the same plane from different points of view. These systems have found practical uses in applications ranging from scientific and medical visualization of complex 3D structures, remote manipulation of robots in hazardous environments, in computer games and marketing [16]. In such systems, the viewer's eyes converge to a point in space in a different plane, which tends to cause fatigue, headaches and eyestrain [5].

In recent years, many 3D recording and display systems have focused on the multi-view technique to create multiple view-windows, where the viewer observes two view-windows to perceive 3D effects. In such techniques, the creation of motion parallax is highly dependent on the change in the multiple view-windows with viewer movement. Such techniques are liable to cause eye strain, fatigue and headaches after a prolonged viewing motion [17].

With recent advances in digital technology, some human factors that result in eye fatigue have been eliminated in the "holography" 3D system. This utilises coherent light sources to construct a true 3D object in space where the image scene is recorded by the interference of the coherent light source. Colour holography produces 3D images of startling quality, with depth cues and parallax that are difficult to achieve in the other techniques. However, a number of challenges need to be overcome before a holographic 3D display is ready for the mass market, due to the high end optical requirements, which means that it is an expensive approach [11].

Despite much work and effort, the currently established methods for acquiring and displaying 3D images, as presented above, still cannot provide a truly realistic 3D viewing experience in an ergonomic and cost effective manner as they all exhibit various drawbacks. A method that could overcome these drawbacks is holoscopic 3D imaging (also referred to as integral imaging), which is a type of auto-stereoscopic 3D display developed from the work of Gabriel Lippmann, 1908 [18]. Holoscopic 3D imaging theory is based on the idea of holographic imaging. Each of the viewer's eyes sees a different picture due to the distance between them, and the brain decides that the object must really be 3D [19]. Recently, technological advances in micro-lens manufacturing as well as increases in processing power and storage capabilities have enabled holoscopic 3D imaging technology to become a practical and promising future candidate technology.

This technique allows 3D images to be viewed without any special eyewear, exhibits continuous motion parallax throughout the viewing zone and presents a variety of different scene views independent on the observers' position with less visual discomfort. Additionally, the holoscopic 3D imaging system is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution through the use of unique optical components. It is also suited for multi-viewer applications with a single aperture camera and conventional flat panel displays, which are equipped with and overlaid with a micro-lens array [21] [22]. These advantages have attracted many interested researchers in the 3D imaging field so that a range of areas such as depth inversing, object depth extraction, electronic display for image compression, optical capturing and post production processes as well as computer generation and refocusing are included in areas being studied presently.

1.2 Research Motivations

Undoubtedly, vision is the most important sense for humans, and different types of 3D imaging technology have endeavored to provide the capability to present remarkable 3D estimation results. Thus, creating 3D content has been the goal of many researchers in academia and industry for many years due to its desired properties. This includes the development of a new generation of 3DTV that can be watched with the naked eye and that produces a 3D effect that is comparable to real life.

Holoscopic 3D imaging (H3DI) technology endeavors to provide a new method of creating and representing 3D images. The principle of this technique uses a naturally occurring process as in the “fly eye” for capturing and displaying 3D images [23]. It is a unique method for creating a true volume spatial optical model of the object scene in the form of a planar intensity distribution through a micro-lens array [24]. It uses natural light and a single aperture camera setup with a micro-lens array in the capture process. Moreover, it offers a full parallax object scene as in the real-world without the need for calibration that is required by other currently available 3D imaging techniques. Therefore, due to these advantages it does not cause eye strain [25]. Holoscopic 3D imaging is a technique that is close to holography, in which a laser beam (as a coherent light source) is used in the capture process; however, this makes holography an expensive technique. Due to progress having been made in the manufacture of micro-lenses, some of the related problems have been solved, for example, limited depth of field [26] and resolution of display images [27]. This has caused the quality of the H3DIs to improve markedly. Therefore, capturing and displaying real-time 3D images via this technique constitutes a promising production 3D technology and has attracted the attention of many researchers. This technique shall be available to the public as a glassless technology [5] [17].

Holoscopic 3D imaging in its simplest form consists of small micro-lenses closely packed together that are in contact with a recording device. When this technology is used, the 3D information is stored in a specially coded 2D format called an "Element Image", in which each micro-lens records one elemental image. As each micro-lens is in-effect a small, low-resolution camera, each of the elemental images is very small. The complete set of all the recorded elemental images constitutes the H3DI in a 2D format. These 3D cues can be replayed with a micro-lens array placed in front of an optical device (e.g. LCD) to reconstruct a true 3D scene. The extraction of the depth information is necessary for the holoscopic 3D image to be processed further (e.g. for coding or feature extraction, etc.). Depending on the type of micro-lens array, H3DI techniques are categorised into two types: unidirectional and omnidirectional parallax.

1.3 Research Aim and Objectives

The research presented in this thesis aims to develop estimation and segmentation techniques that are capable of estimating precise depth information from an observed scene, and are applicable in a range of vision tasks for the H3DI technique. In this investigation, a special focus was given to tasks that favour algorithms with wide generalisation capabilities (no restraints were placed on the algorithms). These have the ability to estimate depth information from both types of H3DIs and features with variable characteristics of objects in the scene (e.g. viewpoint changes, deformable objects, presence of noise and changes in lighting). Indeed, this is required to develop a deeper understanding of the performance of such systems from two directions: the optical model and scene depth estimation perceptions.

In this research four aspects will be investigated for the estimation and evaluation of the depth information map from both real-world and virtual holoscopic 3D cameras, as follows: (1) depth through re-arranging the pixels of the captured H3DIs, (2) 3D feature detectors and descriptors thresholding techniques, (3) 3D object segmentation and (4) H3DI processing including correction and increasing the extracted viewpoint image resolution for depth estimation of the source content.

Recently, a novel and interesting approach for decoding the depth information embedded in a planar recording image was investigated by *Wu, et al*. [23, 24]. In this approach, there exists a correlation between the pixels at a one micro-lens pitch distance interval, and this has been exploited to extract the viewpoint images. A method for depth-through-disparity has been built upon these extracted viewpoint images to estimate the depth information map by exploiting the corresponding displacement among the viewpoint images. However, this approach has the following main limitations: (1) generalization capability, i.e. the approach can only work on unidirectional holoscopic 3D imaging (UH3DI) due to the inherent loss of information associated with local feature detection and description process. (2) Computational complexity (time consuming), where the approach failed to set local invariant features that were well-suited for learning variable 3D objects in the scene, i.e. the approach requires the manual setting of the feature threshold for each captured H3DI. (3) The correct perception of the object scene layout, which depends on whether or not the depth information is realistic or acceptable, i.e. the different planes can be distinguished and their relation in space can be understood. Furthermore, the approach was unable to; (1) Build a 3D object segmentation technique to separate the 3D image into different 3D object planes according to the different object depths, i.e. separate the foreground objects from background (noise). (2) Generate super-resolution viewpoint images to further improve the acquisition of the high resolution depth information map and perspective correctly.

The main objective of this thesis was to provide lightweight techniques for H3DIs that have good depth perception, i.e. algorithms with a trade-off between accuracy and speed. So far, the speed and accuracy of H3DI techniques have been critical for the performance of the whole system.

The new approach presented in this thesis was built upon overcoming the problems associated with the existing multi-baseline disparity analysis algorithm. Significant progress was attained through the development of local invariant features on the objects in the scenes from both processes: setting distinctive features and matching local descriptors to estimate the depth information map through viewpoint image extraction. In addition, the generalisation depth information algorithm was utilised, which is based on the principle of invariability, to produce most of the appearance based variation. Furthermore, the approach can remove the effect of the projection distortions and effect of the camera orientation.

The viewpoint images (VPIs) were formed through re-arrangement of the pixels from different micro-lens H3DIs that are presented in 2D. Each VPI within the H3DI contains a different perspective of the 3D scene and it is within these scenes that the 3D information is contained. However, these extracted VPIs have much higher pixel numbers than the elemental images (EIs) and are still considered to be low resolution images compared to perspective projection images, due to the orthographical projection of the scene. Feature-based match similarity measure algorithms have been used in this approach, and have been widely proven to be the most reliable approach used to estimate the depth map in stereo vision systems [28]. This is a first and crucial step towards obtaining depth information to identify objects in the scene. The goal is to obtain a set of local features that capture the essence of the underlying input images and that encode their interesting structure. A reliable setting feature has been achieved by relating stored geometric model properties and projection parameters from the extracted viewpoint image. Two feature detector and descriptor techniques have been used to improve the performance of the depth estimation related to the generalization, speed and quality.

The depth-through-disparity feature analysis techniques used in this work were: multi-baseline technique, adaptive aggregation cost technique, auto-thresholding feature-based-point detection and description technique, auto-thresholding feature-based-edge detection and description technique, 3D object segmentation techniques, super-resolution viewpoint image generation technique, and depth evaluation technique. In addition, image pre-processing and post-processing techniques were used and implemented on the H3DI for further development and adaptation, including the local histogram analysis technique, quantization technique, smoothing filters technique and cross-validation technique.

1.4 Original Contributions

This work is part of the 3D VIVANT project that is a FP7 specifically targeted research project that aims to capture scenes automatically in 3D space and deliver them for realistic, interactive, fatigue-free and immersive play back to home users/viewers. This work is based on part W4.1 of the project, which is aimed at extracting depth information and obtaining a segmentation of the holoscopic 3D content that can be used for specifying an interaction map. To achieve the above goals the following aspects were investigated experimentally and presented in this thesis:

1. Reviewing

A brief review of the current display technology relevant for 3D. A review of the main H3DI capture and display systems is presented to show the advantages of this technology in terms of simplicity, accurately and cost benefits against other 3D technologies. In other words, it shows that it is a simple and efficient system of capturing real 3D content without having to go through complex dual or multi camera calibrations.

2. Rectification

H3DI rectification is the first and crucial stage to eliminate camera distortions (lens distortion), which cause barrel distortion when sampling all the pixels at the same local position under different micro-lenses. It is important to correctly extract the pixels with the same position in each EI; thus, the H3DI distortion needs to be corrected before extracting the viewpoint images.

3. Setting and Extracting Features

Proposed two novel techniques for setting and extracting invariant distinctive features for further applications. A novel, fully automated optimal thresholding algorithm for setting and extracting feature-point-based on the training VPI has been investigated. To provide more stable features for the illumination variations in noise and wide generalisation capabilities implementation, an adaptive auto-threshold feature-edge-based detector was investigated. This thresholding algorithm enabled automatic extraction of feature-edge maps with reduced noise.

4. Generation of High Resolution View/Multi-view

Proposed a novel technique to set and analyse more feature blocks that can improve the visual quality of the depth map. The novelty of this technique is based on transforming a set of orthographic projection low resolution VPIs into a form of high resolution perspective projection geometry using a modified shift and incorporation method. This proposal is another novel aspect associated directly with the creation of a high resolution viewpoint image. When using the sub-pixels-shift technique for creation of multiple high resolution views, the new generation technique of the 3D multi-view from a H3DI system was investigated.

5. Depth Map Estimation

Proposed an adaptive depth estimation technique for measuring and estimating full parallax depth maps that is, as far as the author knows, the first computer-based technique that has been implemented on omni-directional holoscopic 3D images (OH3DIs). A modified aggregation cost window in the correlation technique was employed to vary the shape of the window across the image depending on local variations in intensity and the majority vote query within the block window.

6. 3D Object Segmentation

Proposed two 3D object segmentation (object extraction) techniques to separate the foreground (object in the scene) from the background (noise) automatically. The object extraction performance has a great importance in achieving the correct object, i.e. identification of 3D object in the scene. The novelty of this work is based on the simplest way of performing the object extraction from the holoscopic 3D content via generation of a 3DIM that combines exploitation of the available depth estimation technique with the 2D segmentation technique.

7. Performance Evaluations of 3D Depth Estimation Techniques

To analyse and evaluate the obtainable results from the proposed techniques, a new form of performance evaluation is investigated for a H3DI system. The novelty of this proposed method lies in designing a specific environment to evaluate the results of a H3DI system where no-reference image for the quality assessment metric is available from this system. This was carried out through the evaluation of the product of each phase of the techniques individually using subjective, objective or combined criteria. Subjective Quality of Experience (QoE) described by the Mean Opinion Scores (MOS) of human opinion was performed to verify the 2D H3DIs' quality, foreground mask quality and depth accuracy. While objective quality metrics to reliably rate the performance were implemented to verify the performance of depth measurement and foreground object segmentation.

1.5 Thesis Outline and Chapters' Summary

The whole thesis is divided into seven main chapters as illustrated in Figure 1.1. The following sections briefly outline the contributions that were made in each chapter and summaries of how the main body of the work was carried out:

Chapter 2- Exploitation of 3D Depth Estimation Techniques

This chapter surveyed and briefly described the main 3D display technologies. The technologies were listed and related to each other using their stereoscopic and auto-stereoscopic 3D depth estimation and display methods. They were classified based on whether or not glasses were required. These technologies are also the ones that are most commonly found being used in 3D television, 3D cinema and 3D games. Then there is a thorough description of the H3DI technique that has been developed by the 3D VIVANT project at Brunel University.

Chapter 3- Holoscopic 3D Depth Estimation Technique Based on Disparity

An intensive review of the most common depth map estimation methods for holoscopic 3D technology is provided. The methodology adopted in carrying out the research is explained and how this technique can be implemented on UH3DIs. In addition, the obstacles and constraints associated with this technique are discussed. Two contributions were made to improve its performance, which were as follows: i) incorporating sub-pixel analysis (up-sampling) computation on the extracted 2D viewpoint images for setting more feature blocks and ii) the implementation of modified aggregation cost windows in the correlation technique using three weighting factors (distance, colour and intensity value) to vary the shape of the window to speed up the process. An additional contribution is the implementation of this adaptive technique on (OH3DIs) to estimate the full parallax depth map, and based on the author's knowledge, this is the first technique to have been implemented on (OH3DIs).

Chapter 4: Hybrid 3D Depth Estimation Techniques

Two automatic threshold techniques were developed for setting and then extracting distinguishing features reliably from orthographic projection viewpoint images. These were used to estimate the speediness and accuracy of depth maps produced from extracted training features models. The detected features were geometrically contained between the multi-baseline viewpoint images in the same scene. The techniques proved to be successful as they detected features easily and solved problems related to image based feature correspondence. Then the two techniques were employed to modify the multi-baseline disparity algorithm so that the performance of the depth-through-disparity method of generating 3D contents was improved. Redundancies in the multi-baseline disparity criteria of the two problems were solved by relating them to the depth in the disparity analysis algorithm via local features. The solved problems were the speediness and invariance of the features that correspond to different multi-baseline images. Both unidirectional and omnidirectional real/computer generated H3DIs were used to test the accuracy of the depth estimation.

Chapter 5- Holoscopic Depth Estimation Based on Super Resolution Viewpoints

Two problems associated with depth through multi-baseline disparity analysis are the accuracy and precision for 3D shape. A longer baseline (more viewpoint images) produces superior precision due to wider triangulation. However, this makes the range of search for the best match larger and leads to an increased possibility of false matches (reducing the accuracy). The challenge in this chapter is to optimize the overall performance of depth-through-disparity analysis in terms of accuracy, smaller number of viewpoint images and less computation, which then leads to an estimate of high-resolution depth maps. Of course this requires additional input features to increase the technique's effectiveness in setting and detecting features to realize the increase in quantity and quality of the reliable feature. This is done to generate super-resolution viewpoint images (SRVPIs) through integrating two compatible techniques i.e. the transformation of a set of orthographic projection VPIs into a perceptive projection image and the deblurring-sharpening filter technique is a necessity to estimate the full parallax 3D depth. The contributions from this chapter are clearly shown in Figure 1.1.

Chapter 6- Holoscopic 3D Image Segmentation

Object segmentation is an important requirement for a wide range of holoscopic 3D depth applications that includes finding the object's borders and separating them from the background or other objects. Consequently the developed segmentation module is a promising topic for research studies. In this chapter two different 3D object segmentation filter techniques based on central (reference) viewpoint image segmentation were investigated and implemented. These filters were implemented from the built-in algorithms and run in the existing framework. They produced the mask through morphological functions that are applied to the center viewpoint image and a threshold equal to the maximum intensity value of the object is then applied to obtain the foreground mask. The mask is produced through seeding and growth around the valid feature sets detected and then extracted from both of the auto-feature setting techniques presented in the previous chapter. The objective and the subjective test scores are used individually in the rest of this chapter to evaluate the performance of; the two foreground object extraction techniques, the 3D Interactive Map (3DIM) creation and SRVPI generated process using the numerical data collected as well as through visual comparisons analysis.

Chapter 7- Conclusions and Future Work

This chapter provides an overview of all the work presented in this thesis and considers how successful it has been. Where hybrid techniques (the integration of two or more techniques) have been developed, they have been highlighted. The effects of the overall impact on improving the depth-through-disparity analysis algorithm are considered. Then finally recommendations are made for future work that will overcome limitations in the technique presented here.

1.6 Author's Publications

A number of journal and conference papers related to this thesis have been published at international conferences and in journals while there are additional papers that have been submitted to international conferences and journals recently. In addition, a list of journal and conference papers that are under preparation is also included.

1.6.1 Published Conference Papers

[1] **E. Alazawi,** J. Cosmas, M.R. Swash, M. Abbod, and O. Abdul Fatah, "3D-Interactive-Depth Generation and Object Segmentation from Holoscopic Image," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, China, June 14-18, 2014.

[2] **E. Alazawi**, M. Abbod, A. Aggoun, M. R. Swash and, O. Abdul Fatah, "Super Depth-Map Rendering by Converting Holoscopic Viewpoint to Perspective Projection," IEEE 3DTV-CON: Vision beyond Depth, AECC, Budapest, Hungary, July 2-4, 2014.

[3] **E. Alazawi**, A. Aggoun, M. Abbod, M. R. Swash and, O. Fatah, "Scene Depth Extraction from Holoscopic Imaging Technology," IEEE 3DTV-CON: Vision beyond Depth, AECC, Aberdeen, Scotland, Oct. 7-9, 2013.

[4] **E. Alazawi**, A. Aggoun, O. Abdul Fatah , M. R. Swash, "Adaptive Depth Map Estimation from 3D Integral Images", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, London, UK, June 2013. "Awarded – Best student paper".

[5] **E. Alazawi**, A. Aggoun, M. Abbod, “3D Holoscopic Imaging Technology for Next Generation 3D Contents,” in Research Student Conference School of Engineering and Design, ResCon13, Brunel University, London, UK, June 2013.

[6] **E. Alazawi**, A. Aggoun, M. Abbod, “Extract Depth Map From Holoscopic Image Using Auto Hybrid Disparity Analysis Algorithm”, in Research Student Conference School of Engineering and Design, ResCon12, Brunel University, London, UK, June 2012.

[7] M. R. Swash, O. Abdul Fatah, **E. Alawazi**, T. Kalganova, and J. Cosmas “Adopting Multiview Pixel Mapping for Enhancing Quality of Holoscopic 3D Scene in Parallax Barriers based Holoscopic 3D Displays,” IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, China June 14-18, 2014.

[8] M. R. Swash, J. C. Jacome, A. Aggoun, O. Abdul Fatah, B. Li, **E. Alawazi**, E. Tsekleves, “Reference Based Holoscopic 3D Camera Aperture Stitching for Widening Overall View Angle ”,10th European Conference on Visual Media Production, UK, Nov. 2013.

[9] M.R. Swash, A. Aggoun, O. Abdul Fatah, B. Li, J. C. Jacome, **E. Alazawi**, E. Tsekleves “Pre-Processing of Holoscopic 3D Image For Autostereoscopic 3D Display”, 5th International Conference on 3D Imaging (IC3D), 2013

[10] M.R. Swash, A. Aggoun, O. Abdul Fatah, **E. Alazawi**, E. Tsekleves, “Distributed Pixel Mapping for Refining Dark Areas in Parallax Barriers Based Holoscopic 3D Displays”, 5th International Conference on 3D Imaging (IC3D), 2013.

[11] M. R. Swash, A. Aggoun, O. Abdul Fatah, B. Li,J. C. Jacome, **E. Alawazi**, E. Tsekleves, “Moiré-Free Full Parallax Holoscopic 3D Display based on Cross-Lenticular”, 3DTV-CON: Vision beyond Depth AECC, Aberdeen, Scotland, 7-9th Oct. 2013.

[12] M. Nawaz, J. Cosmas, A. Adnan, M. Inam Ul Haq, **E. Alazawi**, “Foreground Detection using Background Subtraction with Histogram,” IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, London, UK, June 2013.

[13] O. Abdul Fatah, A. Aggoun, M.R. Swash, **E. Alazawi**, B. Li, J. C. Fernandez, D. Chen, E. Tsekleves, “Generating Stereoscopic 3D From Holoscopic 3D”, 3DTV-Conference: The True Vision-Capture, Transmission and Dispaly of 3D Video (3DTV-CON), pp. 1-3, Aberdeen, UK, 2013.

[14] O. Abdul Fatah , P. Lanigan, A. Aggoun , M. R. Swash, **E. Alazawi**, B. Li, J. C. Jacome, D. Chen, E. Tsekleves, “Three-Dimensional Integral Image Reconstruction Based on Viewpoint Interpolation”, IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, London, UK, June 2013.

[15] O. Abdul Fatah, A. Aggoun, M. Nawaz, J.Cosmas, E.Tsekleves, M. R. Swash, **E. Alazawi**, “Depth mapping of Integral images using a hybrid disparity analysis algorithm”, IEEE International Symposium On Broadband Multimedia Systems and Broadcasting Seoul, Korea 27th Jun 2012.

1.6.2 Journal papers (Submitted)

[1] **E. Alazawi,** M. R. Swash, M. Abbod, J. Cosmas, A. Aggoun, “3D Depthmap Estimation from a Real Holoscopic 3D Image,” IEEE, Transactions on Broadcast. (Submitted)

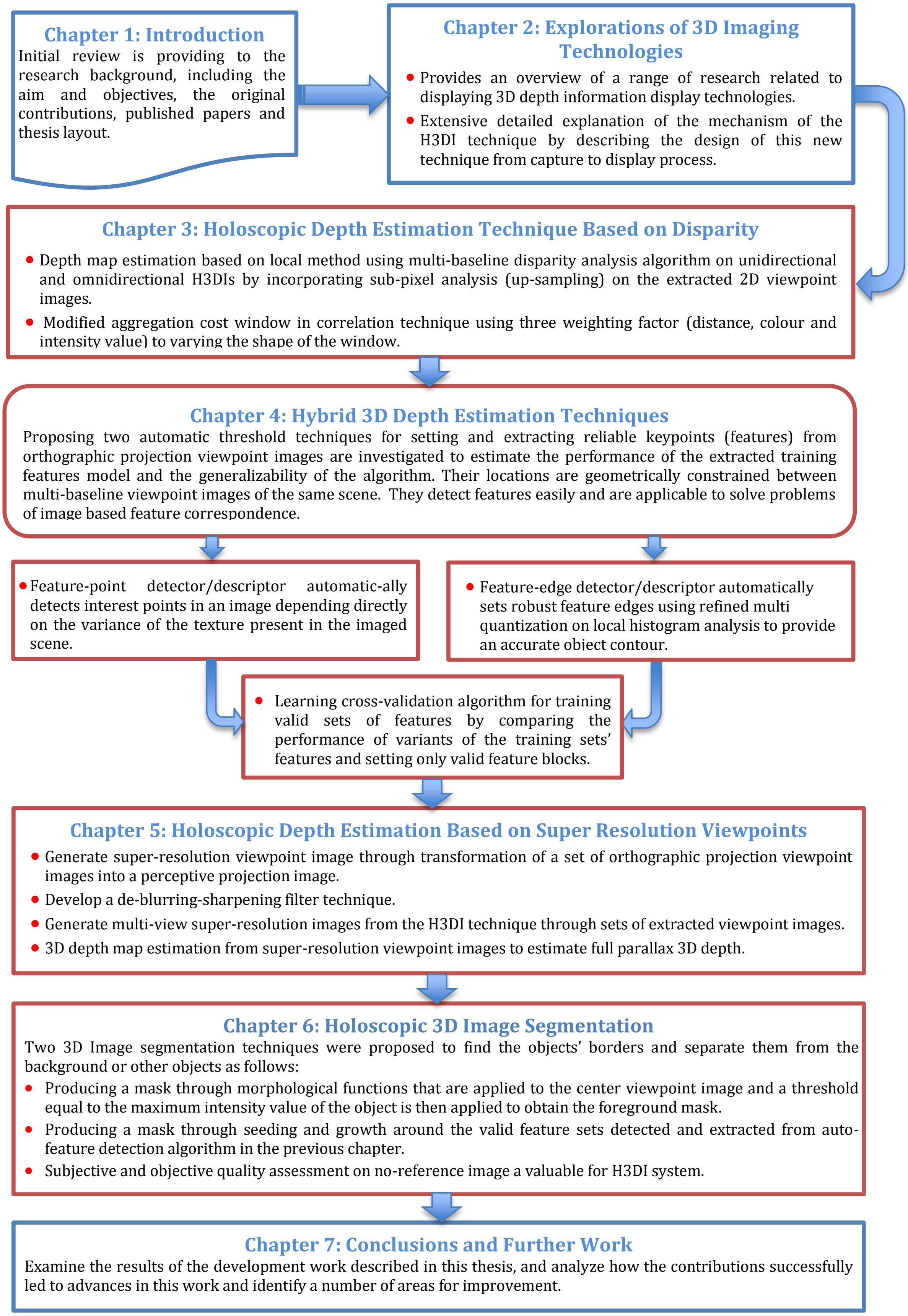
[2] **E. Alazawi,** J. Cosmas, R. Swash, M. Abbod, and A. Aggoun, “3D Depth Measurement froma Holoscopic 3D Image,” IEEE, Transactions on Pattern Analysis and Machin Intelligence. (Submitted Sep. 2014)

[3] **E. Alazawi,** M. Abbod, J. Cosmas, A. Aggoun, R. Swash, O. Abdulfatah**,** “3D Depthmap Creation By Reformatting Holoscopic Viewpoints To Perspective View Images,” IEEE, ACM Transactions on Multimedia Computing, Communications, and Applications (Submitted Nov.2014)

[4] M. R. Swash, O. Abdulfatah, A. Aggoun, **E. Alazawi**, J. C. Fernández, B. Li and E. Tsekleves “Holoscopic 3D Image Re-formatting for Autostereoscopic 3D Displays”, Special Issue on Future of Broadcast Television: Systems, Services and Technologies, IEEE Transactions on Broadcasting 2013 (Submitted Aug. 2013).

[5] M. R. Swash, O. Abdulfatah, A. Aggoun, **E. Alazawi**, J. C. Fernández, B. Li and E. Tsekleves, “Dynamic Hyperlinker for H3D Video Search and Retrieval”. Special Issue on Future of Broadcast Television: Systems, Services and Technologies, IEEE Transactions on Broadcasting 2013 (Submitted Aug. 2013).

[6] O. Abdul Fatah, A. Aggoun, M.R. Swash, **E. Alazawi,** B. Li, J. C. Jacome, E. Tsekleves “Refocusing and all-in-focus image based on single aperture 3D Holoscopic Imaging Camera”, Special Issue on Future of Broadcast Television: Systems, Services and Technologies, IEEE Transactions on Broadcasting. 2013 (Submitted Aug. 2013).



Chapter 1: Introduction

Initial review is providing to the research background, including the aim and objectives, the original contributions, published papers and thesis layout.

Chapter 2: Explorations of 3D Imaging Technologies

- Provides an overview of a range of research related to displaying 3D depth information display technologies.
- Extensive detailed explanation of the mechanism of the H3DI technique by describing the design of this new technique from capture to display process.

Chapter 3: Holoscopic Depth Estimation Technique Based on Disparity

- Depth map estimation based on local method using multi-baseline disparity analysis algorithm on unidirectional and omnidirectional H3DIs by incorporating sub-pixel analysis (up-sampling) on the extracted 2D viewpoint images.
- Modified aggregation cost window in correlation technique using three weighting factor (distance, colour and intensity value) to varying the shape of the window.

Chapter 4: Hybrid 3D Depth Estimation Techniques

Proposing two automatic threshold techniques for setting and extracting reliable keypoints (features) from orthographic projection viewpoint images are investigated to estimate the performance of the extracted training features model and the generalizability of the algorithm. Their locations are geometrically constrained between multi-baseline viewpoint images of the same scene. They detect features easily and are applicable to solve problems of image based feature correspondence.

- Feature-point detector/descriptor automatic-ally detects interest points in an image depending directly on the variance of the texture present in the imaged scene.

- Feature-edge detector/descriptor automatically sets robust feature edges using refined multi quantization on local histogram analysis to provide an accurate object contour.

- Learning cross-validation algorithm for training valid sets of features by comparing the performance of variants of the training sets' features and setting only valid feature blocks.

Chapter 5: Holoscopic Depth Estimation Based on Super Resolution Viewpoints

- Generate super-resolution viewpoint image through transformation of a set of orthographic projection viewpoint images into a perceptive projection image.
- Develop a de-blurring-sharpening filter technique.
- Generate multi-view super-resolution images from the H3DI technique through sets of extracted viewpoint images.
- 3D depth map estimation from super-resolution viewpoint images to estimate full parallax 3D depth.

Chapter 6: Holoscopic 3D Image Segmentation

Two 3D Image segmentation techniques were proposed to find the objects' borders and separate them from the background or other objects as follows:

- Producing a mask through morphological functions that are applied to the center viewpoint image and a threshold equal to the maximum intensity value of the object is then applied to obtain the foreground mask.
- Producing a mask through seeding and growth around the valid feature sets detected and extracted from auto-feature detection algorithm in the previous chapter.
- Subjective and objective quality assessment on no-reference image a valuable for H3DI system.

Chapter 7: Conclusions and Further Work

Examine the results of the development work described in this thesis, and analyze how the contributions successfully led to advances in this work and identify a number of areas for improvement.

Figure 1-1 Illustration of thesis roadmap that shows contributions that have been made in each chapter.

CHAPTER 2

Exploitation of 3D Depth Estimation Techniques

2.1 Introduction

This chapter provides a summary of the main current 3D imaging technologies and 3D depth estimation techniques. It also highlights research into H3DI technology. Therefore the chapter has been split into two parts with the first covering 3D display technology and the second on 3D depth estimation techniques. The first part will include 3D display techniques that require the wearing of special glasses, such as anaglyph, shutter glasses and polarized glasses. Then areas that do not require glasses to work will be discussed, such as holography, volumetric, two-image display, multi-view, light field displays and H3DI. Then finally, H3DI technology will be covered in detail.

The second part will cover 3D depth estimation for spatial change (stereoscopic) as well as spatial and temporal changes (motion). In addition, the second section will include a description of the H3DI system. The rest of this chapter will be an introduction before covering 3D imaging display technology that incorporates stereoscopic 3D imaging and auto-stereoscopic 3D imaging. After that there will be information about the development of H3DI that will cover the history of H3DI, the properties of H3DI and the construction of the camera used in this research. Then 3D depth estimation techniques will be considered including stereoscopic 3D depth maps, motion information techniques and holoscopic 3D depth maps, before a final conclusion.

2.2 3D Vision

Industrial level 3D technology has become much more wide spread in areas such as entertainment and gaming, education and training, data visualization and teleportation, as well as other areas due to the power of computers significantly increasing. Research is needed in the area of using a 2D display to present a 3D object as this is one of the key challenges in the development of 3D-based interfaces. This is where the relationship of objects in 3D space needs to be identified by the viewers (they need to experience depth perception, which is the ability to interpret the world in 3D from visual cues). There are various cues that lead to the perception of depth in a scene, and these cues vary according to the scene in question, the distance to the items being focused on in the scene and the way the information is formed both consciously and subconsciously. Figure 2.1 [29] illustrates the components of the complex and powerful human visual system. The most significant part of this is the depth perception imparted by the two eyes, while other areas are the pre-learned interpretation of 2D images and the unconscious ability to determine motion parallax. The brain merges the information obtained from both eyes to create a sense of 3D [30]. The depth cues from real environments, such as perspective, size difference, occlusion, shadow, accommodation, convergence and disparity, provide a clear picture of the environment that results in an increased performance quality. Thus, it is important to consider that the performance of machine vision systems is also strongly dependent on a good performance of their 3D perception [31]. In these systems, real world 3D effects are replicated as well as motion parallax so the observers can experience 3D effects [32]. This overcomes the four missing cues from 2D images via low-level features and cues [33]:

- Stereo parallax—a different image is seen by each eye.
- Movement parallax—a different image is seen when we move our heads.
- Accommodation—the object of interest is focused on by the eye’s lenses.
- Convergence—object of interest is viewed by both eyes converging on it.

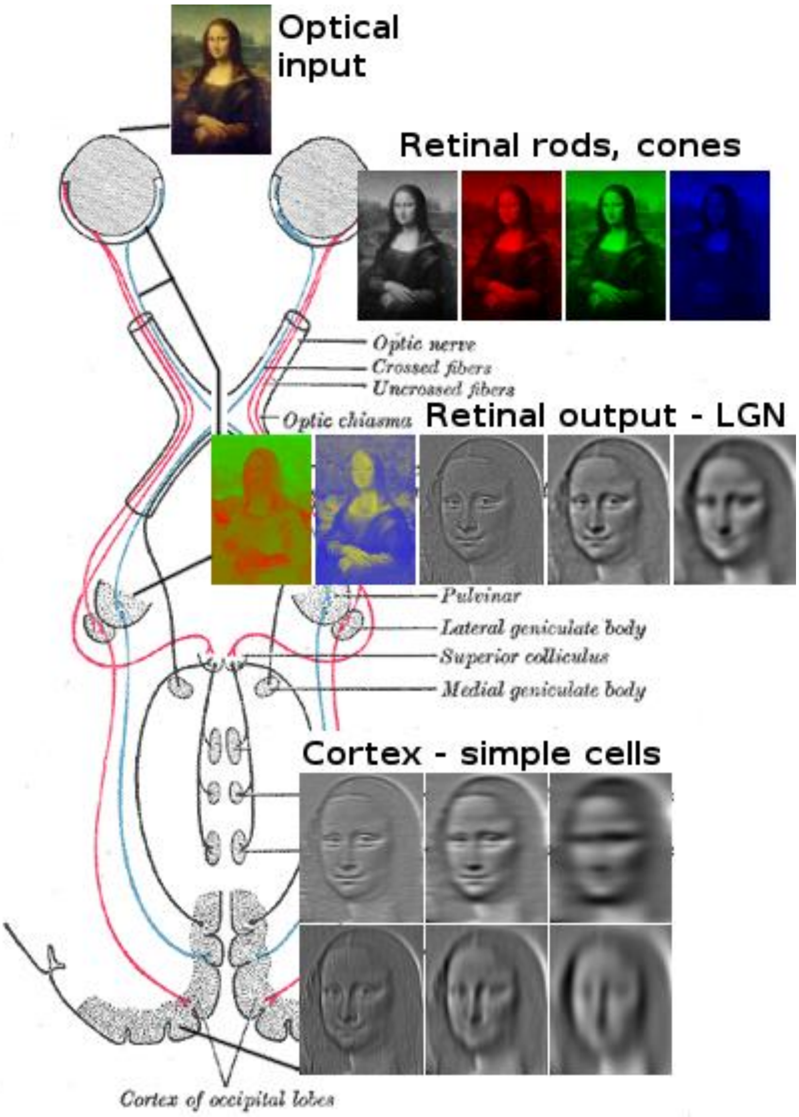


Figure 2-1 Image analysis of Mona Lisa by the Human Visual System (HVS), up to simple cells in visual cortex. By Karel Kulhavy, July 2012 [29].

2.3 Taxonomy of 3D Imaging Display Technology

Figure 2.2 shows the different ways that 3D displays can be categorized from the user's' point of view. This means, the taxonomy for 3D display in this chapter was built upon whether the 3D display technique used by the user requires glasses or not, as a main display category. These techniques are stereoscopic display where optical devices are required close to the viewer’s eyes or auto-stereoscopic display where the eye-addressing techniques are completely integrated into the display itself. The second type is much more technically demanding than the first [34]. The following sections describe the various current 3D display techniques.

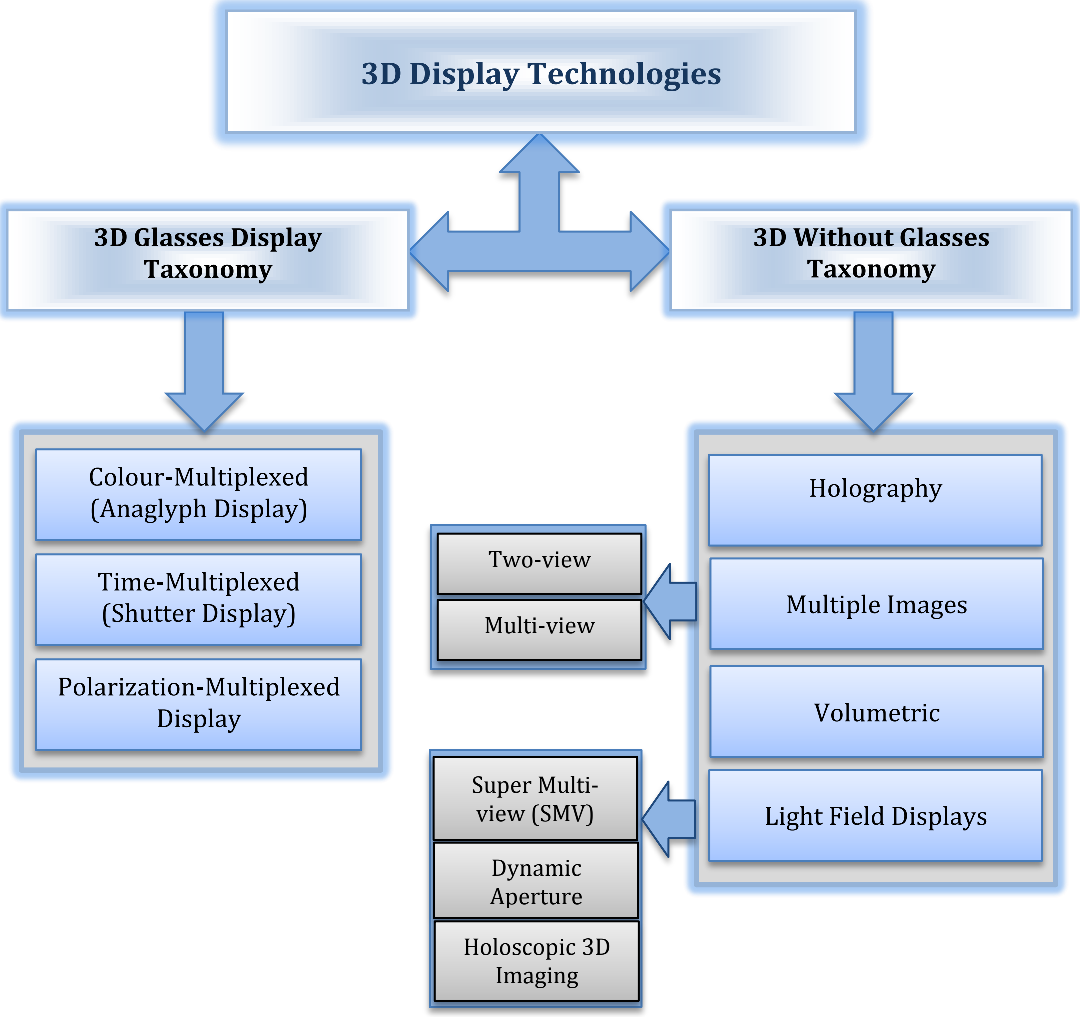

3D Display Technologies
3D Glasses Display Taxonomy
3D Without Glasses Taxonomy
Colour-Multiplexed (Anaglyph Display)
Time-Multiplexed (Shutter Display)
Polarization-Multiplexed Display
Two-view
Multi-view
Super Multi-view (SMV)
Dynamic Aperture
Holoscopic 3D Imaging
Holography
Multiple Images
Volumetric
Light Field Displays

Figure 2-2 Most common state-of-the-art 3D display technologies.

2.3.1 Stereoscopic 3D Imaging Technology

The characteristics of binocular vision are used in stereoscopic images to record 3D visual information that has the illusion of depth as in the real world [35], and the technique was first invented by Sir Charles Wheatstone in 1838 [36] (Figure 2.3). A screen shows two images and a viewer, often bulky, presents the correct image to each eye via spatial or temporal multiplexing [37]. There are three main types of systems for the special displays of the images that can be differentiated via colour, light, time and space multiplexing of the special glasses. The three most used systems will be described in further detail [38].

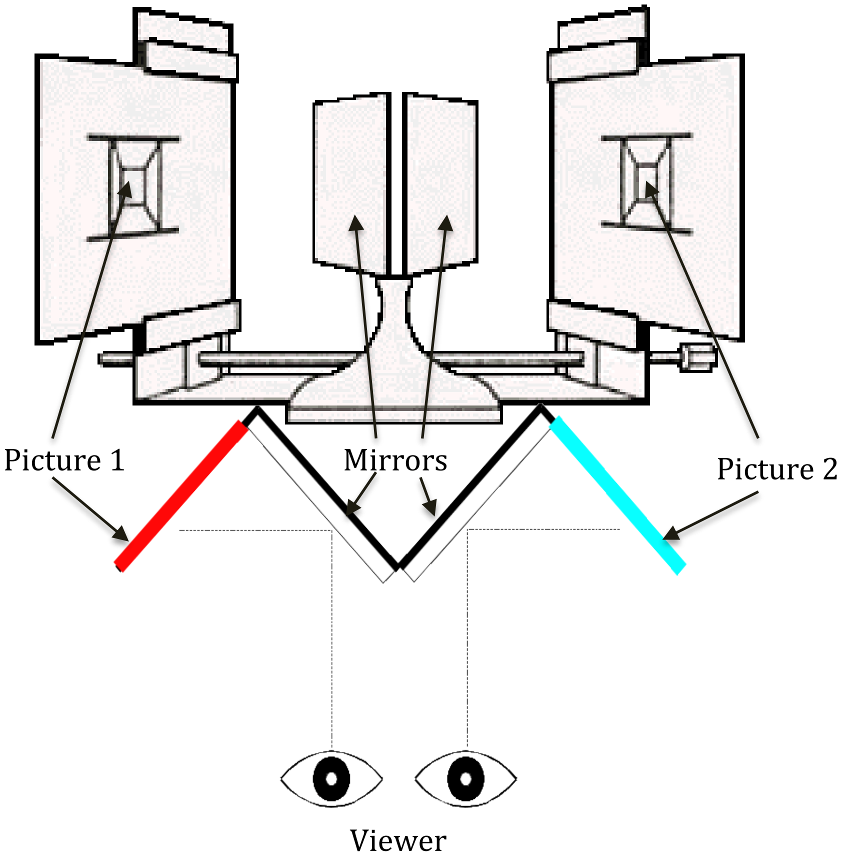


Figure 2-3 Earliest form of Wheatstone's stereoscopic system 1838 [36].

a) Colour-multiplexed (Anaglyph) Displays

Near complementary colours are used in the anaglyph display system to filter the images for the left and right eyes. The possible colour pairs are red and green, red and cyan or green and magenta, which the viewers wear as a pair of colour filter glasses for separation. Limited colour (binocular colour mixture) can be obtained by mixing the red from one eye with the green/blue from the other eye as shown in Figure 2.4 [39]. By using the glasses an image can be viewed that seems to have depth [40]. As human vision is more accurately defined by the dimensions of the L, a, b colour space are L*a*b* values (RGB or XYZ), as they match the lightness, hue and saturation of human colour perception, they have been used in a new algorithm for anaglyph image generation that prevents deficiencies in colour distortions, retinal rivalry and ghosting effect [41].

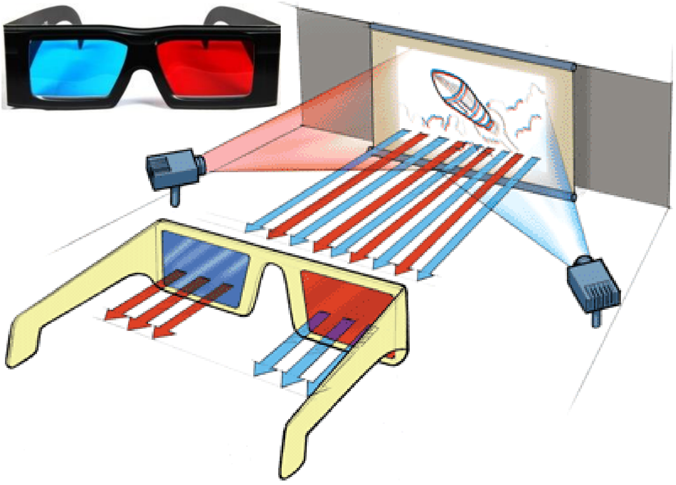


Figure 2-4 Red and blue lens filter glasses and two projected images only, one image enters each eye [39].

This system has the drawbacks of causing eye strain and headaches as well as nausea in some people. In addition, some people have one eye that is dominant and this causes them to not see the image as 3D [42]. As only one image enters each eye, the quality as perceived by the brain is less than with the polarized system.

b) Time-Multiplexed (Shutter) Display

A 3D image can be created via the use of a display screen as well as liquid crystal shutter glasses (LC shutter or active shutter glasses). Shutters in the glasses open and close rapidly to show the left and right eyes alternating images displayed on the screen. The shutter system is controlled by an infrared link and is usually part of a pair of glasses, as shown in Figure 2.5 [43].

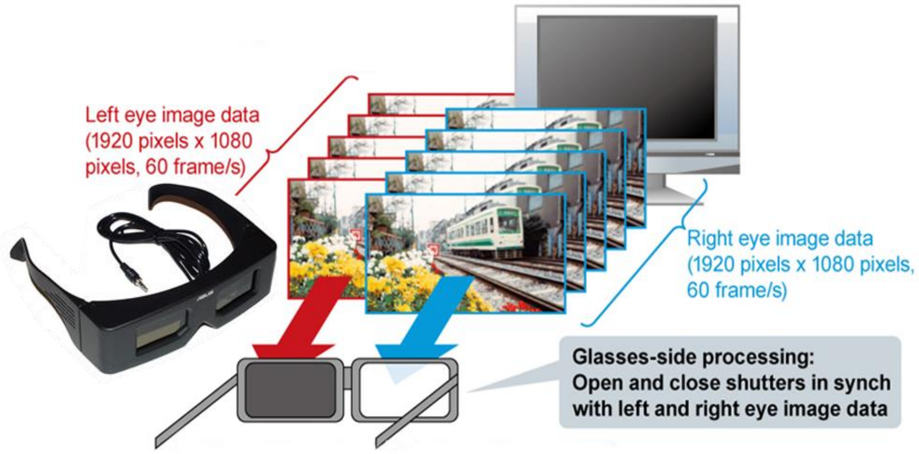


Figure 2-5 Illustrates the principle of operation for frame sequential design and liquid crystal shutter glasses with IR sensor to synch the two [43].

Recently a product has been developed that has rapid switching between the open and closed states as well as high contrast that is a polarizer-based nematic liquid crystal (LC)[44] optical shutters that have the LC cell as the polarization modulator. Despite the recent developments, there are still some disadvantages related to this technique. One of these is that there is a time delay between the two images, which has to be reproduced exactly, as if this does not happen the items in the images will be out of position relatively. In addition, as the image is split between the two eyes, each eye is only seeing half the material, which can result in a notable flicker if the refresh rate is not high enough.

c) Polarization-multiplexed Displays

Currently, the most popular type of 3D system used in the entrainment industry is the polarization technique. In this there are two screens, covered by orthogonally oriented filter sheets (polarization that is either linear or circular), that are positioned so they are at right angles. Linear and circular polarization work in the same way but the use of circular polarization lets the viewers' move their heads without losing the 3D perception [45]. The glasses used in this system filter-out light that has been polarized at specific angles: only light that is compatible with the lens will be allowed to pass through. Therefore, each eye will only see a specific set of images on the screen, as shown in Figure 2.6 [46].

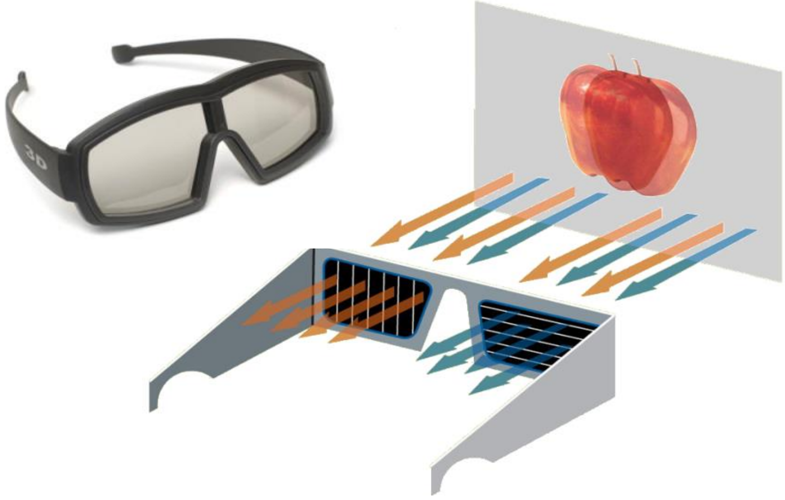


Figure 2-6 Polarized glasses passively block light waves to create the illusion of depth. The lenses in LCD shutter glasses darken when voltage is applied [46].

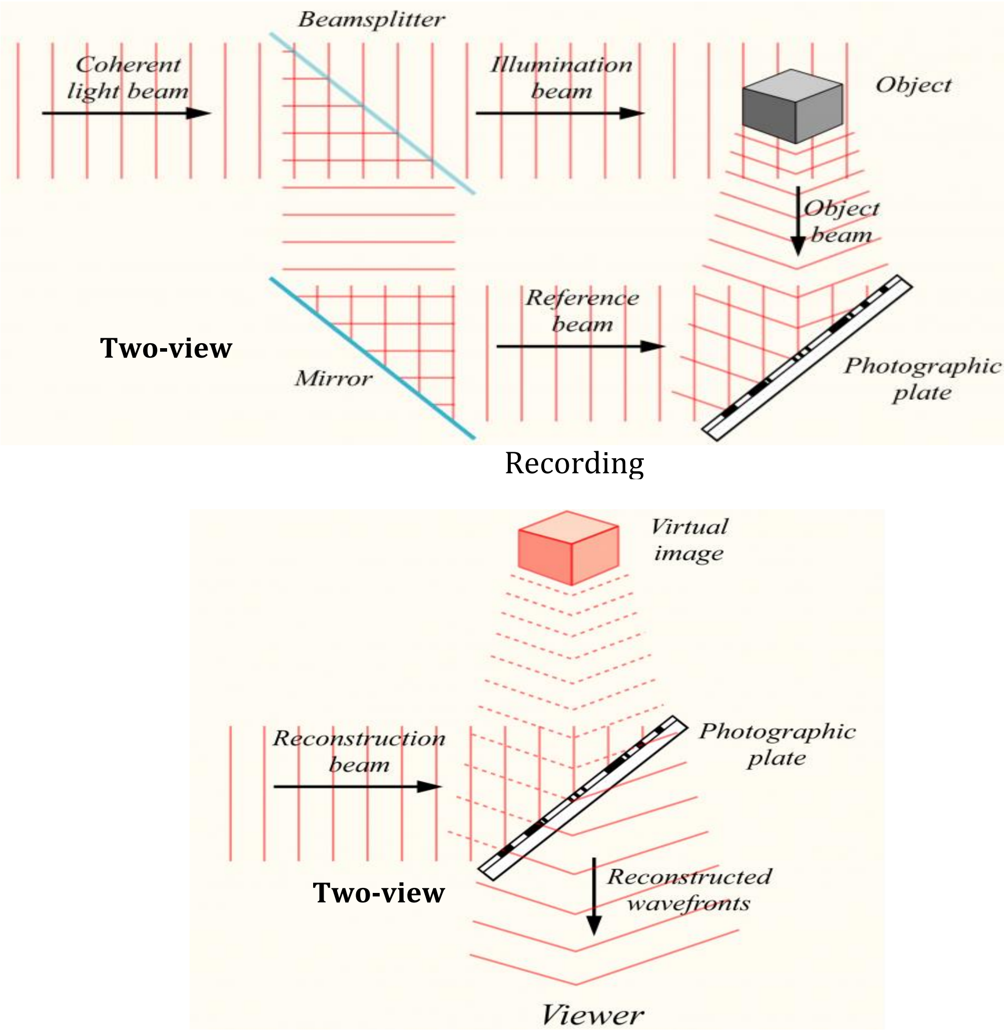
This 3D stereoscopic delivery technique has various disadvantages despite advances in the area of stereoscopic 3D display technology as follows: 1) viewers are required to wear special glasses, 2) viewers may experience headaches and eye stress after extended use, 3) full resolution full natural colour images cannot be delivered to each eye at the same time and 4) there is no motion parallax meaning the 3D scene is only seen from a fixed perspective [47]. Thus, there is motivation for the development of auto-stereoscopic techniques that do not require the use of special glasses.

2.3.2 Auto-stereoscopic 3D Imaging Technology

An auto-stereoscopic display is able to generate a stereoscopic image without the use of glasses, or other devices at the viewer. The divisions within the technology are not always clear but the most useful distinctions divide the auto-stereoscopic displays into four areas as follows: 1) electro-holographic where wave-front reconstruction is used to make the image, 2) volumetric where the image is formed within a volume of space without the use of light interference, 3) multiple images where the viewing field displays many 2D images and 4) a light field display where projector and micro-lens arrays form the image out of light from all directions [48]. The fundamental trends in the newest auto-stereoscopic 3D technology are the areas that this work focuses on.

a) Holography 3D Imaging

Due to them recording and reproducing the light wave-amplitude, wavelength and phase differences, holography or holographic 3D techniques are able to offer 3D depth and motion parallax in all directions, as was proven by Denis Gabor between 1946 and 1951 [49]. However, this could not be put into practice until 1964 when the first hologram was created after the laser was invented in 1960, and then holography was available commercially [50]. Lasers create coherent light that is used in holographic 3D technology to illuminate the scene and camera target when recording and then when the recorded scene is played back again (Figure 2.7). When the hologram is being recorded a laser is divided into two paths with one going to the object being holographed and the other onto the film: the interference pattern between the two beams is the hologram. Then to view the hologram the laser light is reconstructed and sent to the hologram at the same angle as when it was recorded. The hologram diffracts the beam and the image of the object is created [51].



Recording

Viewing

Figure 2-7 Representation of the process for recording a hologram and then viewing the hologram [50].

The need for high resolution emulsions for the holographic recording and reconstruction of films and television has limited its use. However, recently holographic transmission has become digital, and therefore possible, due to developments of CCD capture and spatial light modulators (SLMs). A new holographic projector has been developed at MIT's Media Lab that will enable practical colour 3D holographic video display as well as lower power use when displaying higher resolution 2D images. Guided wave optics was used to build the new projector based on SLMs that are the basis of digital holography. Figure 2.8 shows an example of an image from this projector at 30 frames per second (fps) and a resolution that is comparable to standard definition [51].

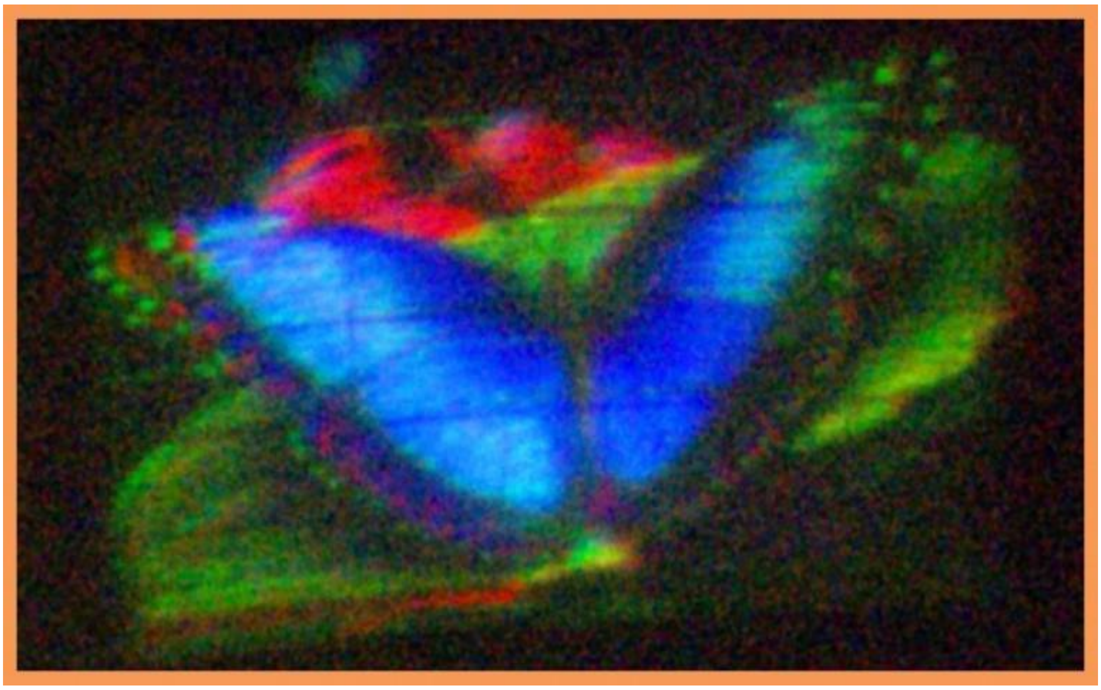


Figure 2-8 MIT's new holographic display showing a butterfly [51].

Due to holography needing coherent illumination and dark conditions when the data is recorded, as well as the images containing a large amount of information, the technique may not be a commercially viable 3D display product in the foreseeable future [52].

b) Volumetric 3D Imaging

Complete parallax is offered by volumetric displays as each point of light comes from a specific point in space [53]. These displays can produce images that are viewable from 360 degrees. These images display accommodation/convergence (AC) and so the focusing of the eyes and the convergence of their visual axes takes place in a harmonious manner [54]. This is also beneficial as the depth cue portfolio can be enhanced by the accommodation. Due to them being composed from a wide range of methods, volumetric displays can be divided into two main types: virtual image and real image [53]. The most common methods use projection screens that rotate, a gas or several liquid crystal panels. Figure 2.9 [55] shows a low-cost volumetric 3D display producing a 360 degree image that was created by the Graphics Lab (ICT) at USC via spinning mirrors, high-speed DLP projectors and very precise computation that calculates the correct axial perspective [54].

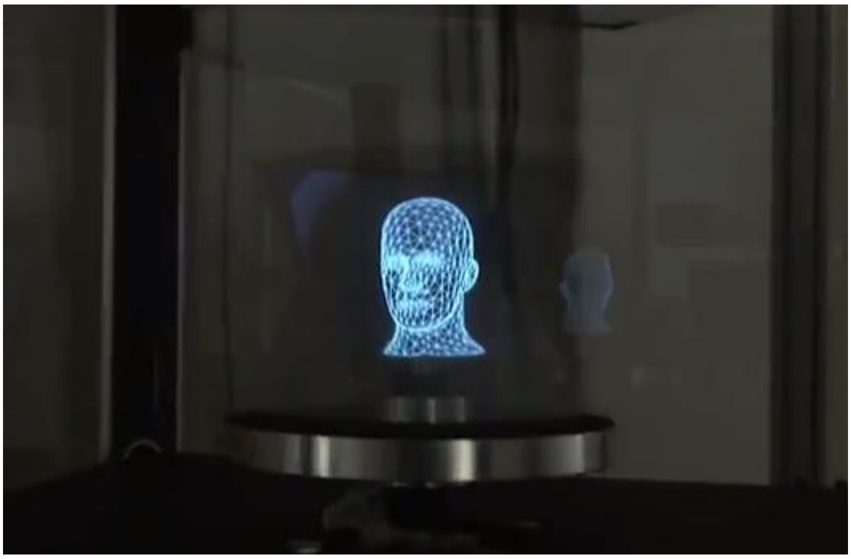


Figure 2-9 Volumetric Display from USC's Graphics Lab (ICT) [55].

c) Multiple Image Display

Multiple image 3D displays can be divided into two systems depending on the amount of views, either two-view display (dual-view) or multi-view display. The two-view display has the left and right views so that they are overlaid as a set of thin vertical strips and displayed at the same time. Then, when the observer is within the 'viewing zone' an optical element is used to make sure that the left view is visible to the left eye and the right view to the right eye. There is a significant difference in the efficiency of the two most widely used approaches at displaying the light on the screen to the viewer from two-view display system [56]. The two common approaches are shown in Figure 2.10 [57].

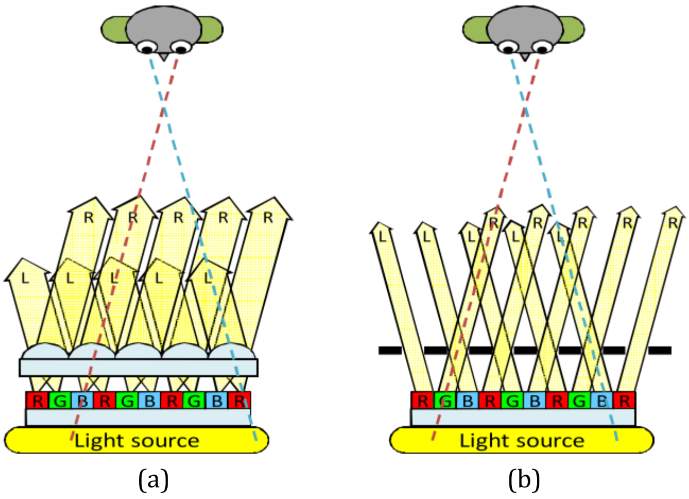


Figure 2-10 Two common methods used in the two-view display (a) lenticular cylindrical lenslet array and (b) parallax barrier array [57].

The first uses alternating cylindrical lenslets in a faceplate and the second has transparent slits incorporated into an opaque surface. Compared to the parallax barrier technology, the lenticular technology has brighter image lighting but it suffers from the moiré effect. These approaches' disadvantage is that pseudoscopic images are formed, that is the viewer must remain within the viewing area for the image to be seen correctly, and if they move away from the ideal viewing distance the chance of seeing the image is significantly reduced [58].

To combat these limitations it has been necessary to add more views, and in these multi-view displays a range of individual views (cameras) are shown over the viewing area. It is necessary to reduce the native resolution of the display in the multi-view display for each zone of the image. This can be achieved by using a vertically-aligned light directing screen as the display that causes the horizontal resolution to be reduced by the number of views. However, the use of a slanted view-directing screen is generally a better option as this causes the transition between views to be made softer so that the viewer experiences continuous motion parallax that is similar to viewing natural images [59]. The multiview display's advantage is that as long as both of the viewer's eyes are in the viewing zone they will experience the 3D image. The multi-view technique typically displays from 8 to 24 views and uses either lenticular sheet or slanted parallax barriers [57] as shown in Figure 2.11. Multiple viewers can be accommodated and each sees the 3D image from their own point of view. To look around an object all the viewer has to do is to move their head [60].

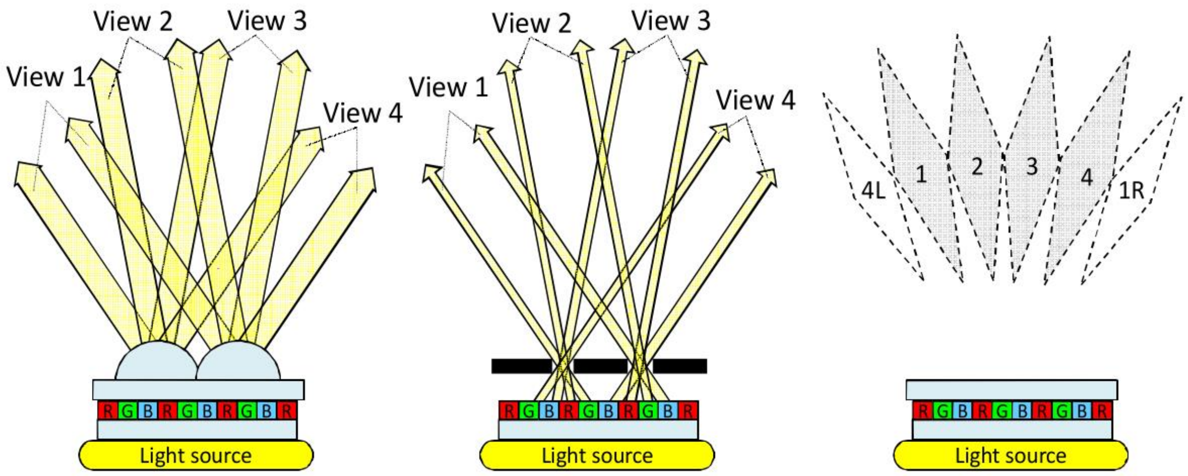


Figure 2-11 Multi-view 3D display using, a) lenticular sheet, b) parallax barrier and c) viewers visibility zones [57].

Development has been taking place on lenticular multi-view displays for over a decade and current performance is limited by the multi-view 3D display because of a reduction on the resolution, the constrained depth of field and the limited viewing regions. As LCDs with higher resolutions are developed the performance in these areas will increase.

d) Light Field Displays

It is possible to capture light field displays in various ways and they aim to create stereoscopic disparity and motion parallax. This means that the viewer does not need to wear special equipment to see a scene as 3D. Generally controlled acquisition is used, such as camera gantries and arrays [61], lenslet arrays or coded aperture techniques, however, it is also possible to use hand-held devices for unstructured acquisition [62]. In this technique, discrete beams of light radiate from each point on the display screen. Here, three fundamental techniques that enable most light field displays will be considered, which are multi-beam, dynamic apertures and integral imaging [63].

I. Super Multi-View (SMV) Displays

A SMV display is a natural 3D auto-stereoscopic display that employs multiple projectors or optical modules to generate 3D images with satisfactory quality. The SMV display technique produces a large number of perspective projections parallax images into corresponding viewpoints. Multi-beam displays optical modules is a type of SMV display where the real image voxels are formed in front of the screen or virtual voxels are formed behind the screen by the multiple beams converging and intersecting or diverging, respectively [63]. The screen diffuses the beams in the vertical direction only allowing viewer’s vertical freedom of movement without altering horizontal beam directions. A new type of light field 3D display is the Holografik technique from the Holografika Technology Company that uses its proprietary light-field technology as the product ‘HoloVizio 360P’ for viewing natural 3D images on its 44” 3D resolution display with 30 megapixels [64] (Figure 2.12).

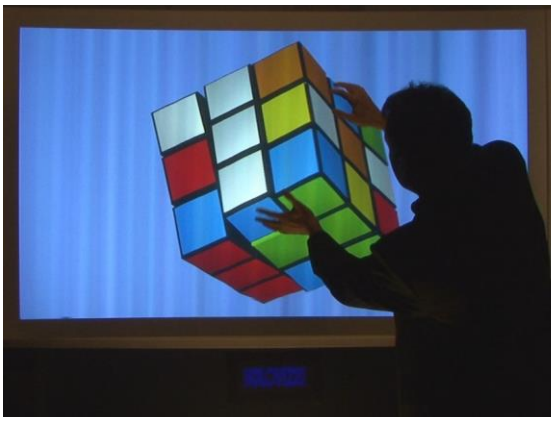


Figure 2-12 Multi-beam light field 3D display using the Holograik technique to produce a natural 3D view [64].

II. Dynamic Aperture

A fast frame rate projector and a horizontally scanned dynamic aperture are used in the dynamic aperture type light field display. When the light passes through the aperture it diffracts in a way that is inversely related to the aperture size, which means that for a spatial resolution the angular resolution is maximised [65]. The visual quality can be further improved as the dynamic aperture can adapt to the content of the image. The results are very similar to those achieved via the multi-beam approach despite the embodiment appearing to be significantly different. The dynamic aperture forms the beams in a temporal manner. The former technology has been developed so that it is able to include head tracking and view steering [66] as well as high-speed temporal modulation [67]. Currently, in combination with high-speed LCD, lenslet arrays are used as rear illumination that can be programed to direct individual views to tracked observers [68]. Multilayer architectures that include a low angular resolution and provide binocular disparity as well as depth cues through motion parallax have been suggested previously [69]. A new "Focus 3D" prototype has been proposed that could generate almost correct accommodative depth cues that would enable the device to be viewed from a wide range of distances [70]. In addition, it would be possible for the camera to focus at a range of depths using this practical 3D prototype. The technique is based on a novel computational display architecture that could support correct accommodation due to its design of the display optics as well as the synthesis of the compressive light field that provides the viewers with almost perfect accommodative and binocular cues as well as motion parallax. Figure 2.13 [71] shows the design of the prototype that is able to let viewers see various depths that are in and out of the screen by sending high resolution light cones to the viewer's eyes.

In light field displays, a new image needs to be presented every 2 *mm* on the screen, which means that a large amount of data needs to be presented to give motion parallax. As there is only horizontal motion parallax the eyes generally attempt to focus on two distances at the same instance. This is an area for further research work.

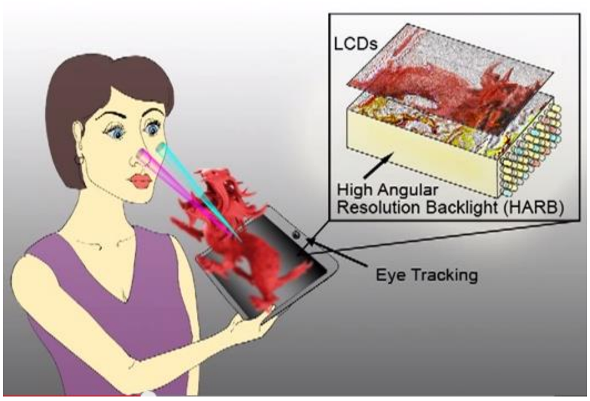


Figure 2-4 Dragon 3D display result focused at two different depths. Using the focus 3D prototype, where a stack of two transparent LCDs is mounted on rails in front of a Fresnel lens with an additional LCD monitor behind the lens the dragon photograph [71].

III. Integral Image Display

The simplest type of light field 3D display technique is integral imaging. This is considered to be a type of holography where the light field is not required to be coherent. The technique records and displays the 3D image as a planar intensity distribution, and it stores the multi-perspective 3D scenes' information in a 2D form [72]. This system was invented in 1908, and since then one of the main problems has been the production of pseudoscopic images, also known as depth inverted images, where the depth is reversed [73] (see next section for more detail). In addition, as the ray bundles spread out, there is very poor depth of field (DOF) in the reconstructed images.

This technique has been developed in the last few decades through improvements that have solved the pseudoscopic problem and enhanced the quality of the lenticular sheet arrays. Therefore, the use of the integral image technique has become attractive and much work has been conducted in this area, for which further details will be given in the next section.

2.4 Holoscopic 3D Imaging Technology

A Holoscopic 3D image display, also referred to as integral images display [72], is a true auto-stereo method type light field display technology that has the ability to provide all four eye mechanism: binocular disparity, motion parallax, accommodation and convergence. This method does not require any special glasses to view the 3D object and does not cause visual fatigue to human eyes. The image is formed by a very large amount of closely packed distinct micro-images. These are viewed via an array of spherical convex lenses.

2.4.1 Holoscopic 3D Imaging: From Past to Present

The use of a series of lenses was proposed by the physicist Prof Gabriel M. Lippmann (1845-1921) on 3rd of March 1908 instead of using opaque barrier lines [73]. This enabled him to record an image with parallax in all directions. An array of small spherical convex lenses known as a fly's-eye lens array was used to record and play back the image (Figure 2.14). The array of lenses is placed over the film sheet or electronic image sensor to capture all the 3D information in a single recording. When the scene is recorded, each lens captures overlapping sections that are recorded in an elemental image (EI) as a special 2D image, and then these are used to form H3DI [72]. However, there were limitations in the depth of field in the method [74].

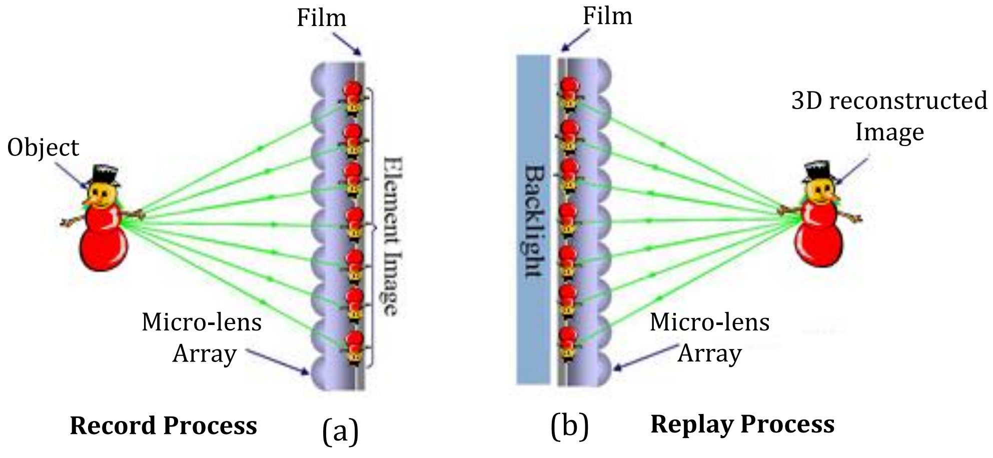


Figure 2-5 Principle of H3D system, a) capture (recording) object process and b) display (reconstruction) of 3D image.

A pinhole sheet was used instead of the plastic materials by Sokolov in 1911 that produced a similar effect but worked on the basis of refraction rather than diffraction [75]. In integral photography (IP), the problem of pseudoscopic images is encountered

(Figure 2.14a) where the image presented is opposite to the orthoscopic image normally seen. The first attempt to solve this problem was made in 1931 by Ives [76] where he suggested recording the reconstructed image a second time and therefore inverting it again. This method, called "two-step integral photography" resulted in a real, undistorted, orthoscopic 3D image being produced. However, the pseudoscopic-orthoscopic conversion problem is not solved by this proposal as the two-step recording process results in image degradation due to the structure of the CCD and LCD that is pixilated and subject to diffraction effects [77]. More noise, distortion and moiré interference are added by the second lens array because of sampling effects, aberration and precision of the manufacturing [78].

The one-step integral solution of generating a pseudoscopic (reversed depth) image to the lens array via computer-generation was suggested by Chutjian and Collier of Bell Labs in 1968 [79] and this produced a correct depth orthoscopic image. The method was improved with the use of colour transparency masks and objective lens cameras by Villums in 1989 [80]. Hain et al. [81] continued to explore the use of diffractive lens arrays for integral imaging. However, there are still the following limitations related to the technique:

- Pseudoscopic image (depth inverse)
- Image resolution.
- Limited viewing angle
- Depth range of 3D scene
- Distortion of 3D scene in adjacent viewing zone

Various companies and researchers considered the mass-production of integral-based products once lenticular products had been successfully mass-produced. The camera system developed by Davies and McCormick between 1998 and 2003 included good developments in the area of H3DI technology that overcame the two-stage recording causing image degradation by using a multiple lens array to optimize the image position and correct the image depth inversions that are caused when the image is replayed through a single-lens array [82, 83]. The correct spatial 3D image can be captured for orthoscopic replay via a two-tier network. The simplest version possible of a two-tier lens array is shown in Figure 2.15 that can make an integral orthoscopic image.

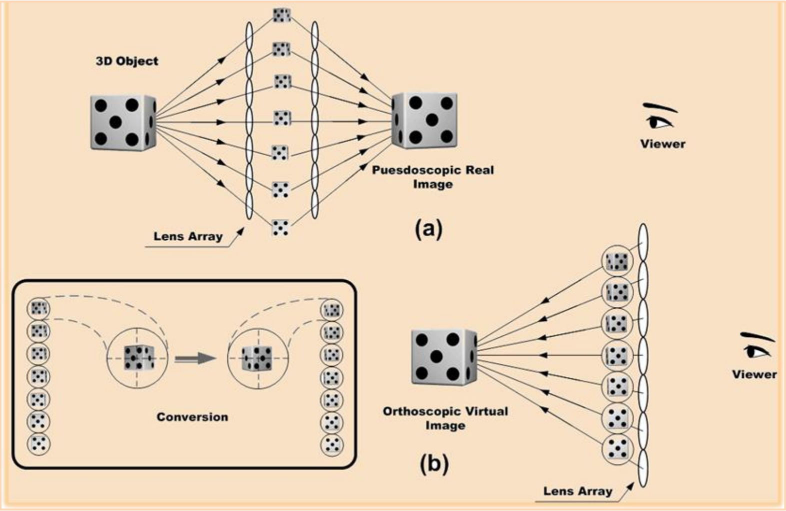


Figure 2-15 Two-step integral imaging for orthoscopic 3D. (a) Pickup and pseudoscopic real image reconstruction, and (b) orthoscopic virtual image reconstruction after EI conversion [78].

The principles of integral photography (IP) have been rediscovered and applied due to advances in micro-lens manufacturing, optoelectronic sensors (e.g. CMOS and CCDs), display devices (e.g. LCDs), increases in processing power and computational storage capacity. Recently, the use of a curved lens array as opposed to a flat lens array has been proposed to expand the viewing angle drastically by a team from Next-Generation Information Display Centre [84]. However, the method fails to keep the space between the lens arrays constant and also has a problem related to overlapping. While the same centre proposed a method where the resolution of the EI is enhanced by jointly estimating the sub-pixel disparity and pixel intensity values from the high resolution grid to be able to use many low resolution EIs to form a high resolution EI [85]. Developments in H3DI techniques obtained from strong research efforts [86] have helped solve problems related to limited depth of field, low resolution, pseudoscopic to orthoscopic conversion, production of 3D images with continuous relief and the observers' limited range of viewing angles.

Very recently, the 3D VIVANT research team at Brunel University has developed new technologies for the capture and display of 3D content for H3DI. A prototype of a 2D camera has been enhanced for the capture of the H3DI. The camera is a single aperture ultra-high definition H3DI camera. For the display, holography is used as it can provide immersive ultra-high resolution 3D content where eye-strain is prevented due to the 3D optical accommodation and convergence working together [72, 73].

2.4.2 Holoscopic 3D Image Properties

The properties of the H3DI come from the geometry of the system used. The concept of the resolution in the H3DI system is not simple. In this process, the lateral and directional information is obtained from an array of lenses and a 2D matrix sensor. This means that the conclusions drawn from either the camera model or display process can be deliberated and transferred to the other due to the high degree of geometrical symmetry. To achieve this, there is a range of properties that are essential for the viewing experience of the 3D scene that need to be understood to determine the resolution of the H3DI systems and they are listed as follows:

- Pseudoscopic image: If the H3DI system is in its standard configuration than the images have reversed depth: they are pseudoscopic (Figure 2.16). The advantage here is that all objects are depth-inverted [78].
- 3D image resolution (3D reconstructed image resolution): This is a combination of spatial lateral resolution and depth resolution. To obtain a measurement of the H3DI system's resolution the number of samples in each lateral plane as well as the amount of rays that cross each other at a sample point should be considered. A point will have a greater differentiation when there are a higher amount of these rays when compared to other points along the depth (axial) direction. This means that the level of error when mapping other points in the H3DI reconstruction is determined by the number of rays in each crossing point. From Figure 2.16, the reconstructed image pixel at depth Z is defined as:

$$\text{Image resolution } (\rho_I) = \rho Z / f \tag{2.1}$$

- Viewing angle: The area within which each EI is displayed (Figure 2.16) limits the angle at which observers can view the reconstructed H3DI, as defined in the following [84]:

$$\text{Viewing angle } (\boldsymbol{\theta}) = 2 \tan^{-1}(P/2f) \tag{2.2}$$

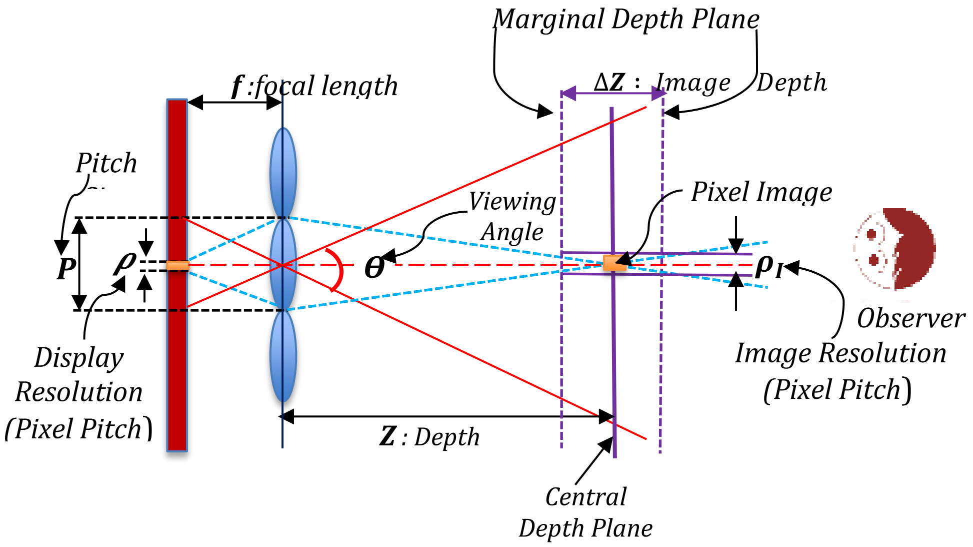


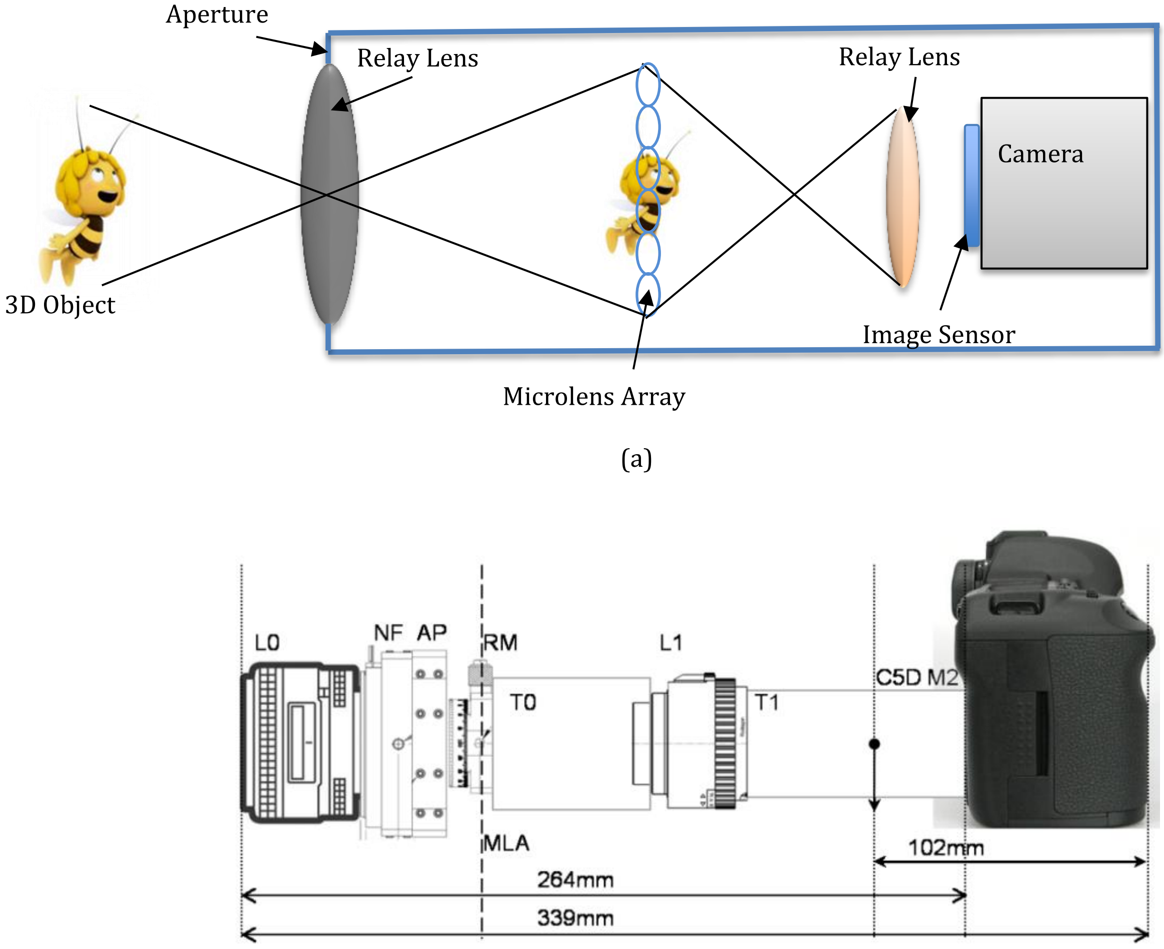
Figure 2-6 Geometrical model of H3DI display, where the various properties of the system are provided and identified.

- Image depth range: If there is an image plane, then objects that are not part of it will have a reduced resolution or be out of focus. The H3DI system's marginal image depth ΔZ represents the range where objects are accurately reconstructed (Figure 2.16) and defined as:

$$\text{Image depth } (\,\Delta Z) = 2(Z\rho_I)/P \qquad (2.3)$$

2.4.3 Holoscopic 3D Camera

The photographic holoscopic 3D camera that has been assembled in the 3D VIVANT project at Brunel University follows the type 2 camera models [87], which has been developed and used to capture H3DIs for processing purposes. The main components of the holoscopic 3D camera are micro-lens array, relay lens and digital camera sensors. The micro-lens array is mounted in close proximity to the sensor as shown in Figure 2.17 (a). The main lens image plane is formed in front of the micro-lens array, which allows the micro-lens array to capture the positions in the scene from different perspectives. The layout for integrating the holoscopic prototype camera with the 5.6k sensor of the Canon 5D Mark2 (C5D M2) DLSR is shown in Figure 2.17 (b). In addition, the total weight of this system is lighter i.e. 1.57 *kg* compared to 3-4 kg for telephotos.



(a)

(b)

Figure 2-17 Assembled holoscopic 3D camera at Brunel University [87].

2.4.4 Principal Components of Holoscopic 3D Imaging

Holoscopic 3D imaging uses lens arrays to collect many views of a 3D scene, and the lens arrays, which are made from hundreds of micro-lenses, can be classified either as a 1D lenticular sheet or 2D micro-lens array. Depending on the application requirements, a holoscopic 3D display can be built on the lens arrays. Depending on the type of lens used, there are two main forms of H3DI. The first is the unidirectional technique, which is a simplified form using a lenticular sheet to form a single direction 3D depth and motion parallax image. While the second type is the omnidirectional technique that uses a 2D micro-lens array that is based on the fly's eye technique to offer an image that has full parallax 3D depth and motion parallax [72, 73]. The structures of both the unidirectional and omnidirectional micro-lenses are shown in Figure 2.18.

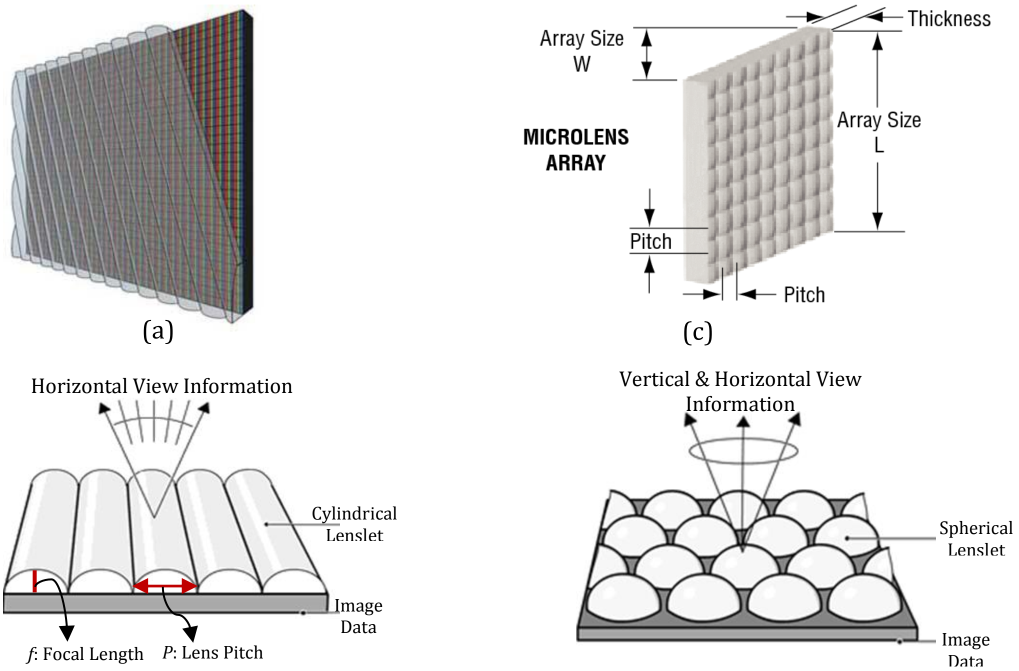


Figure 2-7 Micro-lens structures of H3DI systems [88, 89]. (a, b) Lenticular sheet and (c, b) spherical micro-lenses array.

I. Captured Component Image

It is possible to use only a single sensor and snapshot to capture a 3D scene from a range of different perspectives through the use of a micro-lens array. Each of the separate images is called an EI and each comes from a slightly different view of the scene, and they are recorded with a CCD or CMOS sensor [110]. As each EI has a limited size, the image's resolution is very low. EIs in their matrix are referred to as the holoscopic image of the 3D scene. If an array has K micro-lenses, then K EIs will be produced with L pixels each. Then when the ELs' pixels are rearranged L viewpoint images will be formed with K pixels each [72]. There are differences between the extracted viewpoint images (VPIs) and common images as they are parallel projections: that is they are an orthographic projection. As shown in Figure 2.16, depending on the lens parameters (such as pitch size (P) and focal length (f)) the VPIs are the orthographic projections of the 3D image that have been rotated in a plane by an angle of θ. As shown in Figure 2.19, the 3D scene is being recorded in the VPIs from an angle that is different from that in the perspective projection used in traditional 2D recording.

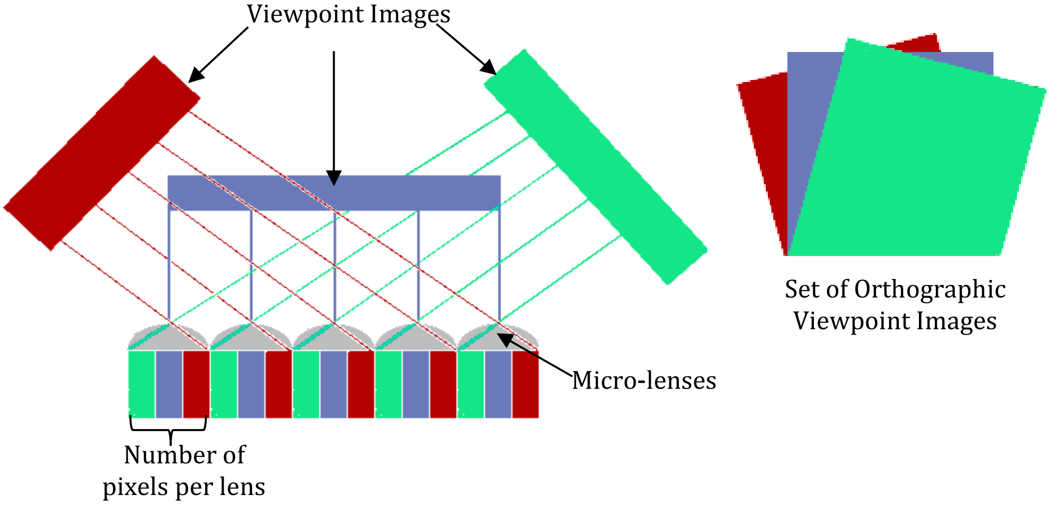

Viewpoint Images
Set of Orthographic Viewpoint Images
Micro-lenses
Number of pixels per lens

Figure 2-19 Shows the orthographic projections for the extracted VPIs (for simplicity, assume there are only three pixels under each micro-lens), i.e. all the pixels in the same position under a different lenticular lens array are represent by the same.

II. Displayed Component Image

To display the H3DI the recording technique is reversed, that is the micro-lens used should be similar. In the holoscopic technology, screens are used to display the 3D images and the light distribution remakes the original scene. Independent of their position, the viewers see the image as 3D. The H3DI properties given in section 2.4.1 need to be considered when searching for the best device for the display of the images. There are some problems to be considered, with the main being the image is pseudoscopic [90]. Then the viewing zone is also defined by the maximum viewing angle. In addition, the lateral resolution can be a problem when only one pixel of the display device is seen via the micro-lens [91] and the viewer sees the same number of pixels as there are micro-lenses. The 3D resolution is generated from the pitch of the micro-lens array and due to this a micro-lens with a very small pitch (P) should be used to build a high resolution holoscopic imaging monitor.

2.5 3D Depth Estimation Techniques

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. It is the simplest and most convenient way used to store and represents the depth cues taken from a scene where the corresponding depth readings (z values) are stored in the array's elements. There are a range of methods to estimate the visualizing 3D depth scenes and they vary immensely in basic principles of recording and display techniques.

There are two systems to generate the perception of depth. In stereoscopy a different image is projected for the left and right eyes and the brain makes the fusion to perceive depth, while in H3DI different light intensities are produced by micro-lenses. Both techniques should be considered in current computer vision applications when developing H3DI data.

2.5.1 Stereoscope 3D Depth Map

In daily life, depth perception is due to each eye seeing a slightly different perspective, and replicating this remains a challenge for 3D technologies. In this section the cues in human visual systems are considered to develop good performance related to accuracy, speed and generalization. By simulating the optical back projection of many 2D images in computers it is possible to develop depth estimation methods. When excluding the pictorial method the most reliable and common approaches can be split into two categories: depth estimation from changes in spatial resolution and spatiotemporal changes [92, 93].

1. Spatial 3D Structure

Stereo vision or spatial 3D structure is the widest used method to obtain information on 3D structure and distance from a scene and the accuracy. It depends on the stereo camera used and the stereo correspondence algorithm, for which much work has been undertaken [93, 94]. The algorithms give either sparse or dense outputs. Sparse outputs come from methods that are based on matching the segments or edges of images; however, if the speed and accuracy of the calculation are increased this disadvantage can be reduced. Using this there are either local or global algorithms. The local methods have increased speed and reduced accuracy, while global methods need more time but have good accuracy. A full discussion on the taxonomy of dense stereo correspondence algorithms has been published previously [94] and an explanation of the depth estimation is available in the next chapter.

Stereo correspondence is very demanding with regards to computational resources, which means that the computation of dense and accurate disparity maps is a problem. Most of the algorithms used can be described using the same structural set [95]. The problem is a long way from being solved and there have even been new restrictions introduced due to advances in related technologies. The reliability of the results is the most important aspect despite real-time frame rates imposing constraints [96].

2. Spatiotemporal 3D Structure

Estimating the 3D structure of a scene from a 2D image stream is one of the most popular computer vision approaches. It is also referred to as Structure from Motion (SfM) or 3D reconstruction from video sequence. Useful information about a scene can be identified from a series of images taken over time by moving cameras by determining the differences between the images due to motion [97]. In the last two decades, SfM based on stereo matching has generated much interest. This is despite it having excessive time consumption when the frame-by-frame approach is applied. The method can also result in flickering-artefacts between frames when the area is highly textured or has fast motion [98]. The artefacts are a disadvantage that might prevent this method's use in 3DTV technologies and teleconferencing. There are also limitations related to its use with real-world video frames [99].

2.5.2 Holoscopic 3D Depth Estimation Techniques

Most of the work that has been conducted into overcoming the problems related to H3DI reconstruction has been related to determining the viewing parameters and enhancing the image generation rather than optical hardware issues. The main problems are limited depth of field [100] and the displayed images' low quality [101]. The areas of research have been related to enhancing the holoscopic 3D image's viewing angle [102, 103] and forming an orthoscopic H3DI [104]. However, there are many data processing issues that need to be addressed to overcome the inherited restrictions and remedy many of the aforementioned problems of H3DIs. These issues require unique computational data simulation and the application of a wide variety of digital image processing. Moreover, digital computers have been used for imaging applications and recent developments in computers allow for the application of digital methods in almost real-time. Depth knowledge or spatial position is one of the area's issues that can benefit a wide range of situations, such as coding and transmission. In addition, depth may also be used to generate virtual image information of a virtual 3D object by combining images from the real-world with those that are computer generated.

The quality of the images produced through computer-based H3DI methods are now achieving better results than those from all-optical H3DI [105]. Through digital image processing it is possible to use Computational Holoscopic Image (CHI) reconstruction and achieve good quality images that are free from diffraction, device limitations and system misalignment [105]. There are many advantages from the ability to determine 3D depth information and generate 3D reconstructions that are related to H3DI applications in areas such as 3D cinema, medical imaging, robotic vision, detection and tracking of people, biometrics and in video games [72, 74]. The following are the recent techniques that have been developed to provide computational simulation of H3DI systems with a micro-lens array related to depth extraction:

a) Depth-through-PSF

The 3D images are reconstructed by extracting pixels periodically from the EI array using a computer. Images viewed from an arbitrary angle can be retrieved by shifting those pixels which are to be extracted. Early work related to the estimation of depth in H3DIs used Point Spread Function (PSF) or intensity distribution function by identifying and matching corresponding intensity distributions in EIs [106]. When the PSF is used to identify the H3DI system the depth estimation is an inverse problem [107]. Due to the loss of information in the direct process used with the model the inverse image problem is ill-posed while the discrete correspondents are ill-conditioned. This means that the method is not suitable for use with real-world H3DI data [108]. Two schemes have been proposed to combat the ill-posedness that enable approximate realistic solutions to be achieved and obtain constrained least-squares solutions as well as give realistic reconstructions [109]. The first was developed from the projected Landweber method [110] and the second is a form of Tikhonov's regularization method [111]. Both of the methods can only be used with simulations based on numerical data [112].

b) Depth-through-Disparity

More recent work has focused on depth-through-disparity approaches as in this the estimation of the relation between depth and disparity is straightforward. First, the horizontal positions of the pixels in the EIs are rearranged to give the horizontal VPIs, and then the disparity field is calculated between the pairs of VPIs. Park *et al.* [113] and Wu *et al.* [112] were the first to report the testing of correlation metrics for disparity estimation. There was also a multi-baseline technique developed to take advantage of the recurrence of information between VPIs. In later work a neighbourhood constriction parameter was developed, due to the depth being piecewise continuous in the space and this enhanced the accuracy of the algorithm [114]. The modified algorithm was validated experimentally using both synthetic and real-world H3DI, and where the background contained noise, non-smooth depth solutions were prepared. In addition, the method needed manual marking of setting and extraction feature points for the estimation of depth maps so it is time consuming and not suitable for use in automatic systems. This can lead to three problems: i) uncertainty and region homogeneity at points where errors are common in the hybrid multi-baseline disparity process, ii) dissimilar displacements related to the object borders and their matching block and iii) 3D object segmentation.

If a set of feature points are first extracted and matched before fitting the surface to the features that have been reconstructed it is possible to overcome the feature reconstruction problem [115]. Two sets of depth extraction parameters were proposed that used sets of “strong” extracted correspondences. The first method considers the depth estimation problem as a 3D optimization problem and uses a H3DI grid and the intersections between a subset of the surface points. However, this method generates non-smooth resolutions due to the non-uniform sampling of the 3D space. The second method needs to merge all depth maps to form the final depth map as it treats the depth estimation as a problem of merging many stereo-like problems. In this method, which is based on the concept of depth-through-disparity, each multiple stereo problem is just made up of a pair of sequential VPIs. Graph cut algorithms were used to determine the disparities between each pair of the image’s pixels. Multiple optimizations are needed to generate accurate depth maps.

A new development has been the use of a block matching algorithm to determine the 3D visualization of a micro-object from the EIs and reconstructed 3D slice images [116]. In this method, optical magnification of the micro-object is performed and a set of 3D slice images are visualized from the EIs via computational reconstruction from the back-propagation. This produces the estimated depth map by matching blocks in the EI and 3D slice image. The sum of absolute difference (SAD) minimizes the errors when using the block matching algorithm. However, the method produces low-resolution depth maps due to the 3D microscopy images having a narrow field of view and the EIs' low resolution.

c) Depth-through- Anchoring Graph Cuts

The methods of Zarpalas *et al.* [115] contain two disadvantages, which are non-smooth resolutions and they suffer from ambiguity in the extraction and matching of the depth map. The extraction of reliable features, called "anchor points", has been suggested as a solution to the previously mentioned problems [116]. The anchor points constrain the optimization procedure, and so this works as an energy optimization algorithm for the depth estimation and this also improves the estimation accuracy and reduces the complexity of the optimization. The method uses graph cuts to estimate the depth of the 3D points rather than the disparities between the pixels in VPIs that are adjacent. The piecewise nature of the depth is combated via a dual energy regularization term for adjacent pixels in the EIs and VPIs [116]. The method produces a smooth scene structure by overcoming the unidirectional H3DIs problems. The method uses a foreground mask to identify the objects from the background. However, the masks could not be produced form the H3DI system as the large non-informative and homogeneous regions make them hard to generate [116]. The problem of a low-resolution depth map has had a solution proposed [117, 118] that uses two sets of slice images. One of the slice images is calculated from the rectangular window (RW) and the triangular window (TW) is used for the other. Block matching is then used to extract the depth information. The block matching error is defined via the sum of absolute difference (SAD).

d) Depth-through-Focused Plane

An object within a plane is focused correctly when it is reconstructed in the correct depth plane. When many plane objects are present it is not possible to measure the total intensity value and determine the depth. When there are multi-objects and an image is reconstructed in the correct plane, noise will be added from the blurred images of other objects that are out of focus in that plane. This means that the depth detection accuracy and quality of the image that is correctly reproduced are reduced. The defocusing depth estimation has been overcome via the use of pixel weight factor (PWF) or the reconstruction of the CHI with depth sensitivity [119]. The PWF is small when a plane image is reconstructed in a plane with the wrong depth, and the pixel can be suppressed. The level of suppression for the wrong depth planes is related to the weighting factor: the level is one if the image is at the correct depth. The unfocused image is suppressed but the focused plane has a lot of noise. This means that the pixel weight method has been used to produce binary foreground object masks. Gaps are removed and edges smoothed in the binary region through the use of the morphological image-closing function. In addition, this is used to combine pixels, eliminate noise and produce separation between the background and the images. The method has been used to detect and reconstruct a clear image that was part of a plane image that was unfocused. However, due to the method having limitations related to object size and the presence of environment noise in the images the suppression of the incorrect depth plane was not obvious.

e) Depth-through-optical flow

Between neighbour frames in motion pictures, the motion vector estimation algorithm is optical flow [120, 121]. One of the crucial issues related to 3D optical information processing from H3DI systems is the limitations in the occluded region, which all the previously mentioned 3D depth estimation and object reconstruction techniques have. In addition, when the disparity based depth extraction methods with pixel-unit accuracy as well as the disparity determined from the depth extraction method, it is only possible to generate discontinuous depth planes and not an accurate depth of the 3D volume, which again is the case in all the pervious methods. This causes an imprecise reconstruction of the occluded object and is referred to as the depth quantization problem. This problem has been eliminated in a proposed method that used optical flow with sub-pixel accuracy for precise depth extraction [122]. In this situation, the optical flow is the pattern of movement of brightness in the scene due to movement capture by the camera [120-122].

The optical flow in the image sequence of the motion picture is either velocities or disparities within the image. Different weights are used to determine the optical flow with high accuracy from the spatial distance, brightness, occlusion state and median filtering. Then the depth is determined in the sub-image array by ensuring the center of the sub-image is the point of reference and that only the *x* or *y*-direction sub-images are the target for the flow calculation. Then the histogram of the center of the sub-image's depth map is used to set the segmentation threshold. To ensure that each sub-image has an accurate depth map the point clouds are sent to the center of the sub-image. From this a triangular mesh reconstruction is used to estimate the point clouds and the occluded region can be reconstructed. Despite this, there are areas with ambiguous depth values and holes that lack depth information in the depth map due to optical flow field errors from occlusions and low-textural regions.

2.6 Summary

The two sections in this chapter have presented a summary of the background studies and the most recent developments. These covered 3D display technology and 3D depth estimation techniques. The literature review covered areas that are related to the work presented in this thesis.

The first part presented developments related to stereoscopic and auto-stereoscopic work. In the stereoscopic area the viewer sees a different image with each eye as they wear a pair of glasses that enable each eye to see a different image. In addition, the method offers stereo parallax. However this method has some disadvantages such as unnatural colour, eye fatigue and motion sickness so is not suitable for extended use. Viewers do not need to wear glasses when the auto-stereoscopic 3D technology is used, and it is based on one of several different concepts. In these methods multiple images are displayed so that the viewer or viewers see a separate image with each eye. However, this means that the viewer needs to be located in a specific spot to fully view the image and there are the drawbacks of eye fatigue and motion sickness. These systems generally use controlled acquisition, which could be camera gantries or arrays, lenslet arrays or coded aperture techniques. It is also possible to successfully use hand-held devices.

Holoscopic 3D imaging is a technique that offers true 3D via the fly's eye technique, and the light field does not need to be coherent. The 3D scene does not have side effects, that is, depending on the micro-lens array, the 3D depth and motion parallax are continuous. Due to steady improvements, this technology has distinct advantages over other methods as follows:

- Full colour real-time 3D images are produced within the viewing angle.
- Horizontal and vertical continuous parallax in both viewing zones.
- Multiple simultaneous viewers.

- Simple optical system (low weight and cost).
- Animation and interactivity in real-time.
- Can be used with large screen formats.
- Post production enhanced through refocusing features.
- Use with both stereoscopic and multi-view 3D image formats.

The second part of the chapter summarized the most common depth estimation techniques for use with computer vision applications. Both stereoscopy and holoscopic depth were considered to gain a full understanding of H3DI data. Work has been conducted on both hardware issues as well as data processing to overcome the restrictions related to H3DIs. One area of significance was the understanding of spatial position (depth estimation), as this is the simplest and most convenient method of obtaining depth information from H3DIsystems.

This promising technique has been limited due to complexity and that a tradeoff is required between the accuracy of the depth estimation and computation time needed. It is a difficult research area to obtain high depth resolution and high-speed computational simulation. More advanced data processing techniques are needed before further progress can be made in this area. The areas that need further work are feature points extraction, feature matching, 3D object segmentation and super-viewpoint image generation.

CHAPTER 3

Holoscopic Depth Estimation Technique Based On Disparity

3.1 Introduction

A depth map is a 2D array where the *x* and *y* distance information is used to store the depth readings. Depth estimation 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 3DTV images. Moreover, digital computers have been used for imaging applications, and recent developments in computers allow for the application of digital methods in almost real-time. There are a range of methods to estimate the 3D depth in a scene with most of them using the stereoscopic principle. The majority of the stereo correspondence algorithms can be described using more or less the same structural set [123, 124]. In stereoscopy a different image is projected for the left and right eyes as shown in Figure 3.1, while in H3DI different light intensities are produced by micro-lenses [23, 24, 125], see Figure 3.2. These techniques were designed based on the human visual depth perception mechanism. There are several factors (referred to as monocular depth cues) such as light and shade, relative size, motion parallax, interposition (partial occlusion), textural gradient and geometric perspective, which help the human visual system perceive the relative distance of objects within a real scene. In fact, depth map estimation techniques try to use these monocular depth cues and imitate the human visual system when estimating the distance between objects to generate binocular parallax (disparity) for the viewer. Due to the unique recording process that exists in the information distribution of the H3DI, common approaches used in stereo vision cannot be directly applied to obtain depth information from the H3DIs.

Previous work by *Wu et al*. [23, 24] attempted to deal with the task by using computationally simulated H3DIs. A novel depth extraction method based on depth-through-disparity analysis has been used. The method achieves acceptable quality results on simulation by using a modified multi-baseline stereo technique. In this method the feature's information points from multiple VPIs are exploited to improve the matching results. Therefore, most of the objects in the scene can be perceived at the correct depth position. However, the method has difficulty to obtain precise depth maps from real H3DIs due to the inherent loss of information associated with the feature selecting and the corresponding process in the multi-baseline disparity algorithm. This can be considered to be an ill-conditioned problem.

The following are limitations of the previous approach [23, 24]:

- Current work on depth extraction has concentrated on unidirectional data involving only 1-D space in the disparity analysis.
- Speed process, i.e. the method is computationally expensive in terms of the number of logic elements required when the operation is being replicated many times to take advantage of the parallel mechanism.
- The algorithm involves using the feature block pre-selection threshold, parameter. This parameter is currently manually operated.
- Sharp/true depth edges i.e. contour 3D object boundary.
- Object segmentation to separate the foreground (objects in the scene) from their background (noise).
- Low-resolution range depth map.

New techniques need to be considered when trying to overcome the limitations stated above when developing a technique that can be used to extract applicable information from realistic real-world and computer generated H3DIs. This chapter presents an improved version of the algorithm in [23, 24] from two aspects: i) improving depth estimation map accuracy by developing a novel an adaptive shape aggregating matching costs (section 3.3.2) over a correlation window, also referred to as adaptive support weight (ASW), and ii) generalizing a previous approach which has been implemented on UH3DIs to OH3DIs (section 3.5).

3.2 Geometrical Analysis Depth-Through-Disparity Systems

3.2.1 Depth Equation for Stereo System

The geometric basis key problem in stereo vision is to find corresponding points in stereo images. Corresponding points are the projections of a single 3D point in the different image spaces. The difference in the position of the corresponding points in their respective images is called disparity (Figure 3.1). Consider two cameras: Left and Right, Optical centres: O_l and O_r. The virtual image plane is a projection of the actual image plane through the optical centre. Baseline (B) is the separation between the optical centres. Scene Point, P, is imaged at P_l and P_r. Since there are similar triangles (P_l, P,P_r) and (O_l, P, O_r):

$$\frac{B+X_l-X_r}{Z-f} = \frac{B}{Z} \Longrightarrow \quad z = \frac{Bf}{X_r - X_l} \tag{3.1}$$

where, Z and f are the depth and focal length, respectively. Disparity (d) is the amount by which the two images of object point (P) are displaced relative to each other. Define, $d = X_r - X_l$, then the depth (Z) is:

$$Z = \frac{Bf}{d} \tag{3.2}$$

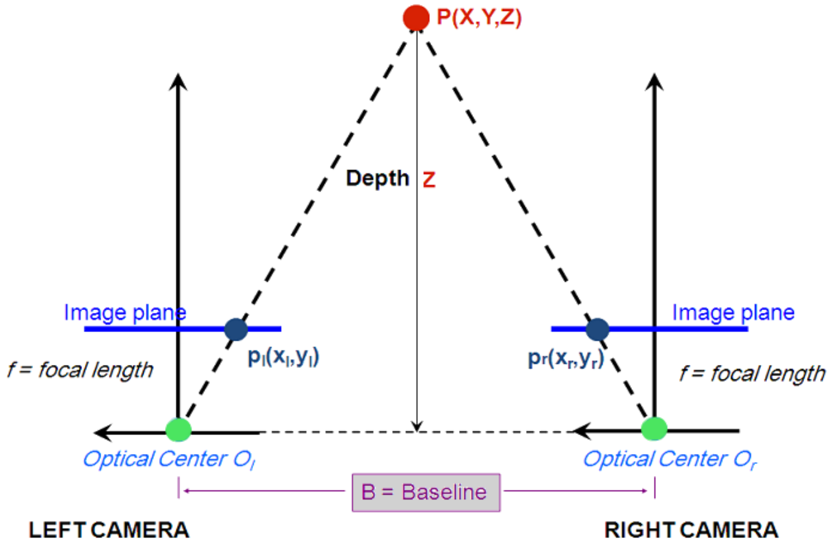


Figure 3-1 Geometric form of a simple stereo vision system.

3.2.2 Depth Equation for Holoscopic 3D Imaging System

As we have seen above, the baseline acts as a magnification factor when measuring disparity from which the depth 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. To determine the method for extracting depth cues from the H3DIs, two aspects are investigated [23-24]: the first is to derive the depth of an object point position through geometrical analysis of the H3DI recording process using the general depth estimation methods used in the computer vision area (e.g. stereo vision). The second area is the mechanism of spatial resolution information encoding within H3DIs. Therefore, the depth equation gives the relationship between the depth and the corresponding displacement between two 2D parallel records of the 3D space images associated with the recording of the H3DI.

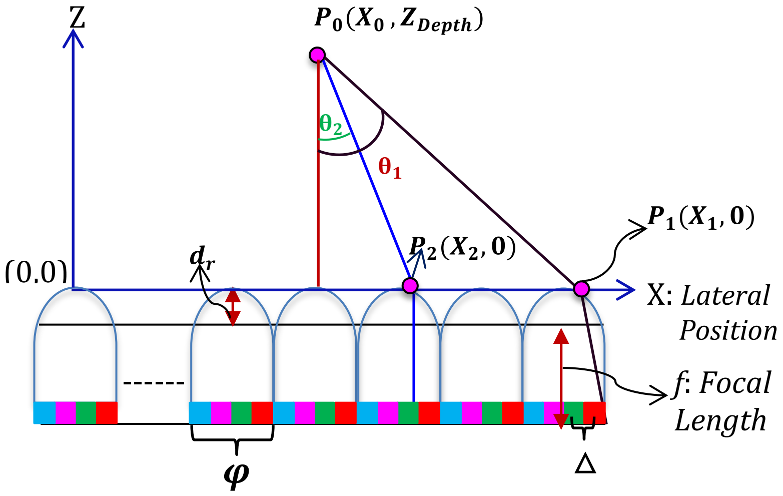


Figure 3-2 Geometric form of 1D lateral direction H3DI system.

To simplify the analysis of an object point $P_0(X_0, Z)$ we will just consider one lateral direction (*x*-direction), though the analysis is easily expandable to the two-dimensional lateral plane. As shown in Figure (3.2) the matrix sensor behind each micro-lens recorded an EI, which was obtained from a slightly different point of view. 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. From Figure (3.2) the mathematical relationship for depth and disparity can be simplified to:

$$Z = \frac{d \cdot \varphi . f}{\Delta} \tag{3.3}$$

where, $\Delta = ds_1 - ds_2$ is the sampling distance between two adjacent pixels, φ is the pitch size of the micro-lens sheet and d is the disparity of the object point $P_0(X_0, Z)$ within two resampled pixels from two EIs i.e. one micro-lens is the disparity unit distance between two resampling pixels in the original H3DI data. A detailed explanation on resampling all pixels in the H3DI at the same local position under different micro-lenses is presented in the following sections. As seen from Eq. (3.3), the range for the depth estimation from two resampling EIs is inversely proportional to the baseline, which indicates that a longer baseline can give a better accuracy. Therefore, ambiguity exists when determining the exact position of the object point from two resampling EIs, which can be reduced by using the information from another resampled EI.

3.3 Multi-baseline Disparity Map Algorithm

Stereo matching uses the multiple-baseline stereo pairs algorithm that was first developed by Okutomi and Kandade [126] to compute the disparity of multiple stereo image pairs with different baselines. The matching is performed by accumulating the evaluation function from each stereo image pair and then making a judgment from the accumulated evaluation function. The judgment is carried out on the score function from each pair directly. The final result is obtained from all intermediate judgments. The criterion to measure similarity of matching features between two stereo images is the sum of squared differences (SSD). The SSSD (sum of SSD) is the intensity value decisions about the correspondences made from multiple stereo pairs. The score function has been made over a window that is the simplest and most efficient criteria [127]. In this approach, the multiple VPIs can be viewed as being similar to multiple stereo image pairs used in the multi-baseline stereo. However, the VPIs are orthographic projections (low resolution) and stereo vision images are perspective projections (high resolution). In addition, the baseline in the VPIs is the distance between two cylindrical micro-lenses, which is different from the distance between the two cameras used in stereo vision.

Accordingly, the extracted VPIs are often used to track the displacement of individual features from VPI to VPI. The selected desirable features to track are often obtained in the same way as proper features to match them. In the matching process the small displacements from pairs of VPIs (i.e. reference VPI and target VPI) have been calculated using the sum of square difference (SSD) cost function. The implementation of feature points on the reference VPI has led to an increase in the stability of 3D solutions. The disparity displacement is extracted, which is the physical distance between corresponding pixels in pairs of VPIs. To determine the initial disparity of the central pixel p within the window block, the SSD is used as a score of the cost match function for disparity estimation and is mathematically described as:

$$C(p,d) = f\ (I_1\,(p), I_2\,(p+d)) \tag{3.4}$$

where, $C(p,d)$ is the cost function, d is a horizontal displacement vector between a pixel (p) in I_1 reference VPI and $(p+d)$ in I_2 target VPI. To determine the initial disparity of the central pixel p (x_1, y_1), the score for all neighbouring pixels (bn) is calculated using different values of d ($d = 0,1,....,d_{max}$); then the value of $\vec{d}_p^*$ is chosen that gives the pixel the minimal cost among all computed disparity cost functions of (bp). Hence, $\vec{d}_p^*$ has the minimum given by:

$$\vec{d}_p^* = \arg\{\min_{d\in R}\{\ score\ (bp, \vec{d}\ )\}\} \tag{3.5}$$

where, bp represents the window around pixel (p) and R is the research area. The cost match function for the initial disparity map estimation $\forall\, d\ in\ x\ direction$ is mathematically described as:

$$C(p,d) = SSD(d) = \sum_{x,y\in w} [\hat{I}_1(p) - \hat{I}_2(p+d)]^2 \tag{3.6}$$

where, $C(p,d)$ is the cost function, $\hat{I}(p) = I(p) - \bar{I}$ and $\bar{I} = \frac{1}{N}\sum_{p\in w} I(p)$, N is the number of pixels within the window (w) and d is a horizontal displacement vector between pixel p over the window (block) w in the I_1 centre VPI and $(p+d)$ in the I_2 target VPI.

The Multi-Baseline Algorithm (MBA) matching score function is used where there are a number of stereo pairs (K) to estimate the initial disparity map. The initial matching decision is carried out using the built up score function from all the VPIs pairs with different baselines:

$$SSSD(d) = \sum_{i=1}^{K} SSD_i\,(d) \tag{3.7}$$

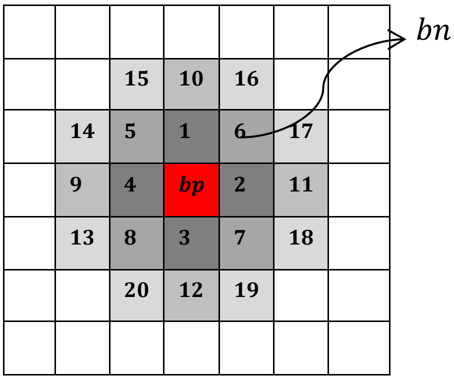
$$C(bp,d)_{initial} = SSSD(d) = \sum_{i=2}^{K} \sum_{p\in w} [\hat{I}_1(p) - \hat{I}_i(p+d)]^2 \qquad (3.8)$$

where, SSSD is the sum of SSD which minimizes the depth map (Z) and *i* is the target VPI (*i = 2,3,……K+1*). Disparity is estimated in the 'winner-takes-all' method by only selecting the disparity label with the lowest cost [23, 24]. The disparity is verified whether it is dominant in the block centered on pixel p (bp). If it is, then this disparity is considered as the final disparity and no further refinement is required; if it is not, further refinement is required. The disparity can be more robustly estimated if the disparity within the neighbourhood is considered in the H3DIs multi-baseline algorithm. *Wu et al.* [23, 24], addressed the problem of choosing an appropriate matching window size in the disparity analysis and neighbourhood constraint and relaxation technique. To improve the matching position of a feature block, the neighbouring blocks are considered rather than individually considering each block.

The matching position of a feature block and the number of neighbour blocks has been used to allow local variation of the disparity in considering the fact that the expected disparity of a neighbour block is not essentially equal to the centre block. The matching decision was carried out using the below built up score function from all the VPI pairs with different baselines and the result is a data file that contains the corresponding disparity between VPI pairs. Considering the neighbourhood relaxation, the score function can be written as:

$$score\ (bp,d) = SSSD\ (bp,d) + \textstyle\sum_{bn\in N\ (bp)} w\ (bn\ ,bp) \times \min_{\delta}\{\mathrm{SSSD}\ (bn\ ,d +\ \ \delta) \qquad (3.9)$$

where, bp represents the window around pixel(p), whose disparity is to be determined, N is a set of neighbouring blocks of center block bp and bn is the neighbour block as shown in Figure 3.3. The parameter δ has been used effectively to allow some depth variations between the neighbour blocks. The score function has the minimum disparity as $\vec{d}_p^*$ and $SSSD\ (bp,d)$ is the sum of SSD that minimizes the disparity of the centre feature block bp. The weighting factor $w\ (bn,bp)$ between bp and bn is introduced to reduce the estimation error caused by setting extra emphasis on the center block rather than the surrounding neighbouring blocks.



		15	10	16		
	14	5	1	6	17	
	9	4	*bp*	2	11	
	13	8	3	7	18	
		20	12	19		

Figure 3-3 Centre feature block $\boldsymbol{bp}$, neighbour block and the sequences of neighbour blocks ($\boldsymbol{N}(\boldsymbol{bp})$) to the centre block.

A higher weighting factor is assigned to a block which has a great possibility of being in the same object as the feature block. Only the distance of colour space was used from *Wu et al.* [23, 24], in which the distance factor (DF) is defined according to the inverse distance to the central block. Those blocks that have high colour similarity to the central block will have a relatively high weighting factor.

3.3.1 Setting and Matching Feature Blocks

Local features are points on the 3D object that are considered as interesting or salient. The significant features from each image must be extracted to identify the matching features in a set of images, and the number of features is related to the image's spatial resolution. *Wu et al.* [23, 24] considered the block's variance before the disparity analysis to set the number of feature blocks for the matching VPI pairs. A block will be a low confidence block (one that is invalid) when its intensity variance is less than a determined threshold. The only blocks that are retained are those with the required number of features to give a confident match. The threshold varied between images and is set manually. If the threshold is increased then more information is needed resulting in a greater number of blocks being classified as invalid. For each image the feature block selection threshold (FBST) is manually determined. There can be missing depth cues on the final depth map due to many factors between VPI pairs, e.g. noise, lack of texture, occlusion, and unaligned surfaces. This means the depth with the lowest cost and therefore may not be the true depth. Due to the manual aspect of this procedure it is very time consuming, and therefore it cannot be used in real-time situations. This means that further work is needed to develop an algorithm that can deal with these issues and provide a good result.

3.3.2 Adaptive Aggregation Cost Window

Local algorithms reduce ambiguity by aggregating matching costs over a correlation window [127]. The correlation window also refers to the local support region that implicitly indicates that the depth is equal for all pixels inside. This intrinsic assumption will lead to numerous errors, especially at regions of depth discontinuities [128]. The support from a neighbouring pixel is valid only if such a pixel has the same disparity. The way to select appropriate support is a key factor of the correlation method [129].

The main problem in depth based multi-baseline methods is how to match the flexibly selected shape and size of the matching window that are based on the content of the VPI pair [130]. To improve the performance of the algorithm, an adaptive weighting factor (AWF) algorithm is used for refinement of the disparity map to prevent multiple neighbouring blocks from being selected as interesting for the same feature and to obtain large support areas in untextured zones, while adapting shapes near the object boundaries. Moreover, unreliable pixels are properly extrapolated from surrounding high confidence pixels. This is done by setting extra emphasis on the center block than the surrounding neighbouring block. The weighting factor set by the correct neighbourhood can be flexibly adjusted to match the center window in the process that contributes to the scoring function. The development is based on computing a confidence kernel distance weighting factor to detect decisive regions with ambiguities (homogeneities, similarities, etc.) that might cause mismatching (untrackable region). An adaptive *k-NN* (nearest neighbors) weighting factor filter has been applied to the initial disparity values center block bp. The weighting factor $w\ (bn, bp)$ between bp and bn is used to increase the frequency of the initial disparity d_p. This can be made dependent on spatial distance to the central block and the confidence of the neighbouring blocks where a block with high variance always contains more information. The adaptive weighting factor is a combination of three confidence terms: $conf\ (w_p)$ = $\exp(-\left\|I_{bp} - \overline{I_{bn}}\right\|/\sigma^2)$ that represents the high variance of the center block *bp* and neighbouring blocks *bn,* where, I_{bp} and $\overline{I_{bn}}$ represent the center pixel value *bp* and average value of the pixels within the block *bn,* respectively. Using a small δ allows for a reduction in the variance within each block. The distance term $dist(w_p)$ =$\sqrt{(i_{bp} - i_{bn})^2 + (j_{bp} - j_{bn})^2}$ is calculated as the spatial Euclidean distance between the coordinates where the high variance is closer to the center block *bp*. The colour difference term is calculated as the Euclidean distance between neighbour blocks in the CIELab colour space of pixel *bp*(C_{bp}=[L_{bp},a_{bp},b_{bp}]) and pixel *bn*($C_{bn=-}$[L_{bn},a_{bn},b_{bn}]), marked as: $c_{diff(w_p)} = \sqrt{(L_{bp} - L_{bn})^2 + (a_{bp} - a_{bn})^2 + (b_{bp} - b_{bn})^2}$. The dual distance-weighted *k-NN* (*k*-nearest neighbours) majority vote method was employed to select the neighbours' size *k* of the *bp* block to reduce the weight of other neighbouring blocks outside the found *k* nearest neighbour [131]. Thus, this lead to the reduction of the sensitivity in the choice of nearest neighbour blocks in the corresponding process. If *k* is very small, the local estimates bias will be very poor. This enables us to develop a robust and precise filtering smooth cost aggregation function by summing the SSD functions of the windows in the neighbourhood of the disparity score function. The aim of all of the above is to produce good performance in feature matching and the corresponding process and accelerate the searching strategy. By this strategy, the source region is limited to a smaller region surrounding the centre block. Different weights are given to the distances of the neighbours to set the corresponding dual weights function to the centre block. Thus, the closer neighbouring block is weighted more. The steps are as follows:

1. Firstly, a set of k with high variance labelled target neighbours for the centre block b_p is identified where a block with high variance always contains more information.
2. Compute the distance D between the centre block b_p and other neighbours b_i, where $D = [d(b_{p,} b_1), d(b_p, b_2), \dots, d(b_p, b_N)]$.
3. Let $F = \{(b_i^N\ , K_i^N)\}_{i=1}^k$ denote the set of the *k-NN* to the centre block b_p arranged in an increasing order in terms of Euclidean distance $D = (b_p, b_i^N)$ between b_p and neighbour block b_i^N ∀ $i = 1,2,3, \dots \dots, N$. Where, b_i^N is the *i*-th neighbours of the centre block b_p and K_i^N is the label for the *i*-th nearest neighbour among its *k-NN*.
4. Arrange the distances in D in an ascending order and store in $D_m = [d_{m1,} d_{m2}, \dots, d_{mN}]$ where d_m is used to represent the ordered distance, e.g., d_{m1} is the distance between the first nearest neighbour and the center block b_p.

5. Let $W = \{w_1, w_2, \dots\dots, w_k\}$ be a set of the corresponding dual weights. By assign to the $i - th$ neighbours b_i^N of the centre block b_p a dual weight w_i is defined by:

$$w_i = \begin{cases} \frac{d_{mk}-d_{mi}}{d_{mk}-d_{m1}} \times \frac{d_{mk}+d_{m1}}{d_{mk}+d_{mi}} & , \ if \quad d_{mk} \neq d_{m1} \\ 1 & , \ if \quad d_{mk} = d_{m1} \end{cases} \tag{3.10}$$

6. Then the distance-weighted factor between b_p and neighbour blocks b_i^Ncan be calculated by applying majority weighted voting (K), where, γ as the Dirac delta function takes a value of one if $\gamma(K = K_i)$ and zero otherwise.

$$w_p \ = arg\, max_K \sum_{(L_i^N, K_i^N) \in F} w_i \times \gamma \begin{cases} 1 & if \quad K = K_i \\ 0 & if \quad K \neq K_i \end{cases} \tag{3.11}$$

The aggregated cost is computed as a weighted factor of the bp for the different neighbouring blocks and can be written as:

$$W(bn, bp) = conf\ (w_p).\ dist(w_p).\ c\_diff(w_p) \tag{3.12}$$

The voting scheme used was from the pixels in the support region; therefore, each pixel p collects votes from reliable neighbours as:

$$\text{Vote}_{p}\,(d) = \{n | n \in U(p) \wedge d_n = d\} \tag{3.13}$$

where, n is a consistent pixel from a reliable neighbour in the support aggregation window of the centre pixel p and the disparity d_n contributes one vote, which is accumulated in the set Vote p (d). The final disparity of the bp is decided by the maximum majority weighted vote number and defined as:

$$d_p^* = arg\, max\left|Vote_p\ (d)\right| \tag{3.14}$$

The dual distance-weighted *k-NN* is an effective method for reassembling neighbouring hard problems that search for and allocate independently a set in the neighbouring blocks N of (bp). This function is employed to reduce the influence of the sensitivity from two aspects: data sparseness and mislabelled blocks (noisy, ambiguous) of the nearest neighbour blocks. Furthermore, it produces good performance in the matching process by recovering the boundary for the objects (untrackable region) and reconstructed scenes with high spatial smoothness. Moreover, its power-smoothnessface (rising when two neighbouring blocks are allocated to the disparities) is inserted into the original feature block bp to produce an efficient total disparity score.

According to Eq. (3.9) and Eq. (3.12), the total cost match function in Eq. (4) resolves into, $C(bp,d)_{Out} = C(bp,d)_{initial} + C_{smooth})$, where $C(bp,d)_{Out}$ represents the filtered output disparity at the bp block, while $C_{initial}$ and C_{smooth} indicate the cost match function of feature block bp and the nearest neighbour block (bn), respectively. The adaptive filter weighting factor $W(bp,bn)$ is applied to the initial disparity $C(bp,d)_{initial}$ at center block bp to increase the frequency of a particular initial disparity before a standard median operation. Finally, the cost score disparity map is written as:

$$C(bp,d)_{Out} = C(bp,d)_{initial} + \sum_{bn \in N(bp)} W(bn,bp) \times min\{C(bn,d+\delta)\} \qquad (3.16)$$

The depth is achieved directly by implementing all the extracted VPIs at the location where the cost score match function has the minimum cost. The right side columns of Figure 3.7 and Figure 3.9 show that the optimal performance results are achieved. The post-processing takes place to recover other blocks by applying the dual distance-weighted function over *k-NN* block.

3.4 Depth Estimation from UH3DI

Only horizontal directional parallax is contained in the UH3DI as the capture of the image only uses a 1D cylindrical micro-lens array. At a fixed-time, both intensity and directional 2D information can be stored in the holoscopic 3D camera. All the 3D spatial information is contained in a 2D format and it is therefore called an EI. In a H3DI, each EI contains information obtained from slightly different directions. The number of EIs is related to the number of lenses that record the 3D object from different perspectives. The recording contains the depth information embedded in a unique way. Along the outline path of the image, an image profile analysis is used to determine the intensity values and then along the horizontal line an auto-correlation is conducted. In an H3DI the pixels' intensities are closer to those from a micro-lens distance than those that are adjacent.

3.4.1 Transform UH3DEIs into Viewpoint Images

The key idea behind the EI transform step is the resampling of the collected data into the form of a 2D image for the subsequent process. This resampled image is called the viewpoint image (VPI) that contains information of the scene from one particular view direction. To obtain a single orthographic VPI view from the UH3DI data will require the resampling of all the pixels in the same location under different micro-lens (e.g. one particular pixel from all micro-lens images at once), resulting in a one view direction of the recorded scene. Resampling is the mathematical technique used to generate a new image with different sizes (height/width) and resolutions in the same image. For instance, if there are *N* cylindrical lenses and each micro-lens covers five pixels, thereby offering *N* EIs and extracting five orthographic, as is clearly illustrated in Figure 3.4.

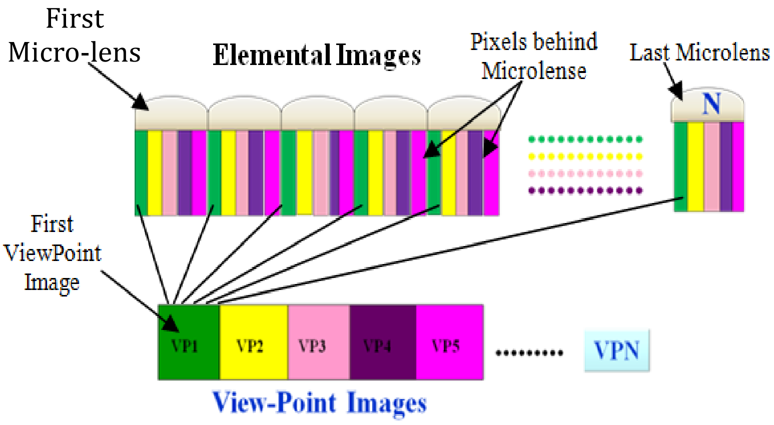


Figure 3-4 Principle of reconstruction of VPIs from UH3DI by periodically extracting pixels from the captured EI (for simplicity, assume there are only five pixels under each micro-lens), i.e. pixels in the same position under different micro-lenses, represented by the same pattern, are employed to form one viewpoint image.

The extracted VPIs are orthographic projections, which is a type of parallel projection with rays perpendicular to the projection plane as shown in Figure 2.19 in Chapter two. Hence, the intensity matching between these VPIs is much easier when compared to EIs. The resolution of each VPI is equal to the resolution of the micro-lens array (or sheet), therefore, it produces low resolution images.

3.4.2 Experimental Results

This section presents the results of the above described approach on two data sets. The algorithm has been implemented on the real-world and synthetic datasets UH3DIs [132] using MATLAB software. To evaluate the above approach and compare it with the presented work in this study, the same parameters have been used in the test including different given thresholds for setting and extracting matching feature blocks.

I. Synthetics UH3DI Database

To test the algorithm, three various types of computer generation UH3DIs were used for the implementation of the above approach. These were noiseless test images, in particular designed for the matching task, therefore it was essential to add some noise to test the effectiveness of the algorithm. The noise can significantly degrade the accuracy of the automated computational saliency detection algorithms; meanwhile, it provides more stable results and implicitly suppresses the noise during the process of computing [133]. The noise added to the test image was a white Gaussian noise with variance σ = 0.5. Figure 3.5 shows the synthetic H3DIs and their extracted VPI, where each image portrays a multiple object 3D scene. In the first and second images (bed and toy, respectively) the 3D object scenes had a diverged texture quality and were transformed to fit the Field Of View (FOV) of the computer-generated micro-lens array. The virtual camera and the complete dataset are publicly available at [132]. The micro-lens array consisted of 99 cylindrical lenses with a 7 × 700 pixels resolution; therefore, the resolution for the extracted VPI = 7 × 700 pixels, lens pitch size = 0.5 mm, focal length $f = 4\ mm$, lens height pitch = 50 mm with a resolution of 700 pixels and the resolution for each H3DI size = 693 × 700 pixels [131, 116]. The third computer generated H3DI (CG-balls) used for the test contained several objects at different depths and a plain background, and it had pitch size (ψ) = 0.6 mm, focal length f = 1.237 mm and lens height pitch = 0.88 mm. The parameters listed in Table 3.1 were used in the algorithm to estimate the final depth map, and were considered and confirmed as being suitable parameters previously [23, 24].

Table 3-1: Main parameters used to estimate final depth map from synthetic UH3DI.

Main Parameters	Ball	Toy	Bed
Number of VPI	7	7	7
Block size	10	10	10
Research area (R)	3	3	3
Neighbouring blocks number (N)	8	12	12
Neighbouring weighting factor	0.5	0.8	0.5
Extracting features blocks threshold	1	1.5	1.2

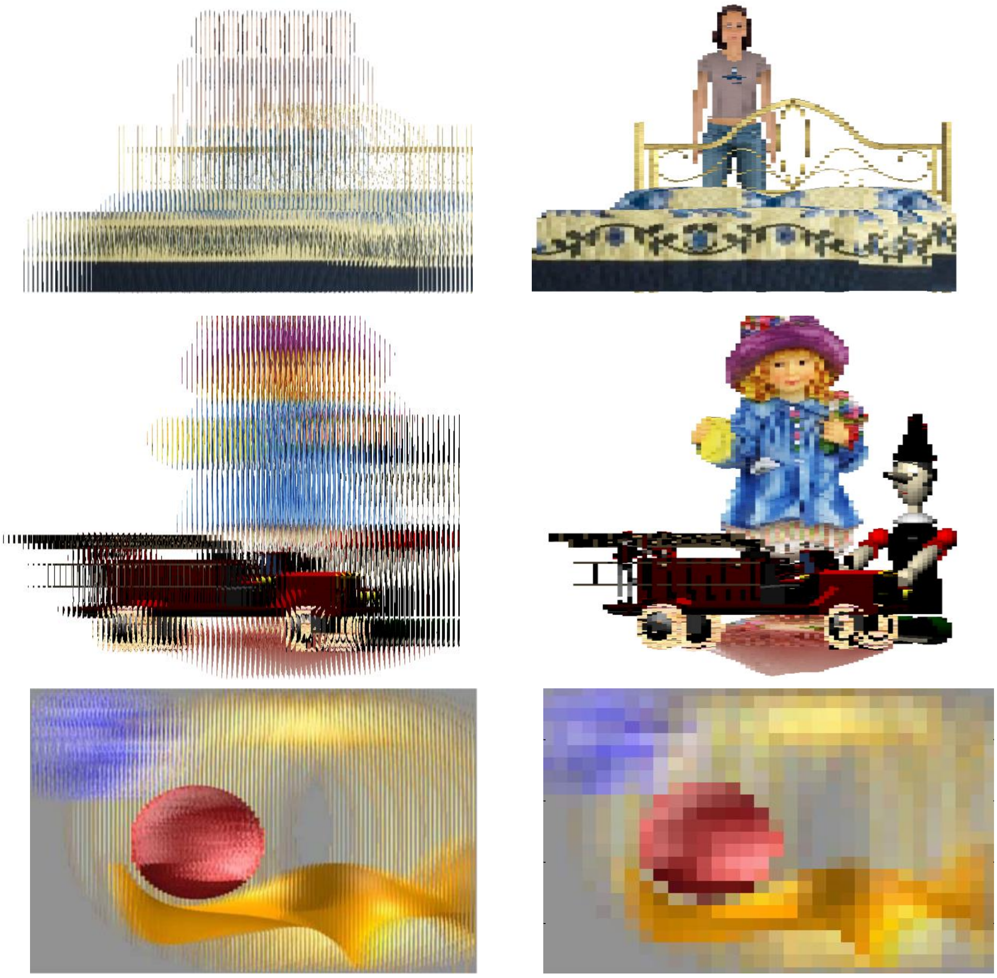
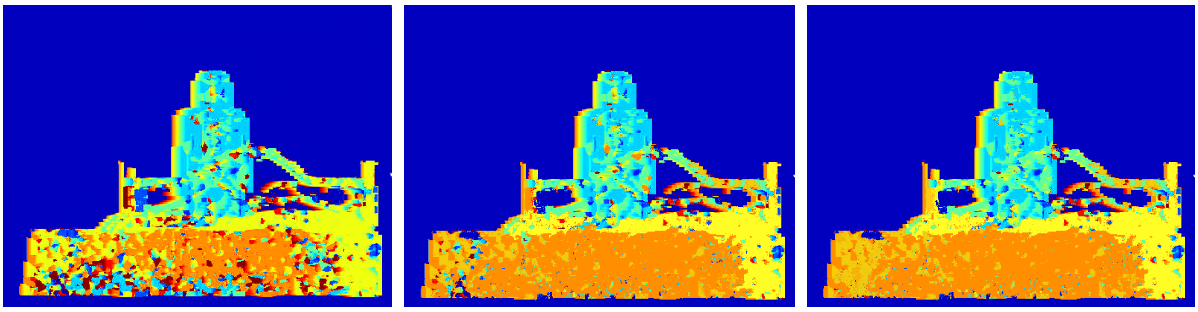


Figure 3-5 Left column computer generated UH3DIs and right column their extracted VPIs.

Figure 3.6 shows and compares the depth maps obtained from different feature block threshold parameters without using post-processing to recover the invalid blocks and smoothing the final depth map. The red colour indicates, the near object to the camera, while the dark blue is the background of the scene and colour calibrator between red and dark blue is the depth planes of the objects in the scene. It is seen that the high feature block threshold value can increase the untraceable position, i.e. one block of pixels may look identical to another block of pixels and consequently more than one position has been detected. This is because it requires more feature information for the matching window. Figure 3.7 shows the corresponding depth map obtained from [23, 24] (left side column in Figure 3.7) and the further improvement in the performance of the depth map (right side column in Figure 3.7) obtained after employing the adaptive shape and size support window to recover the invalid blocks from the surrounding neighbouring blocks.

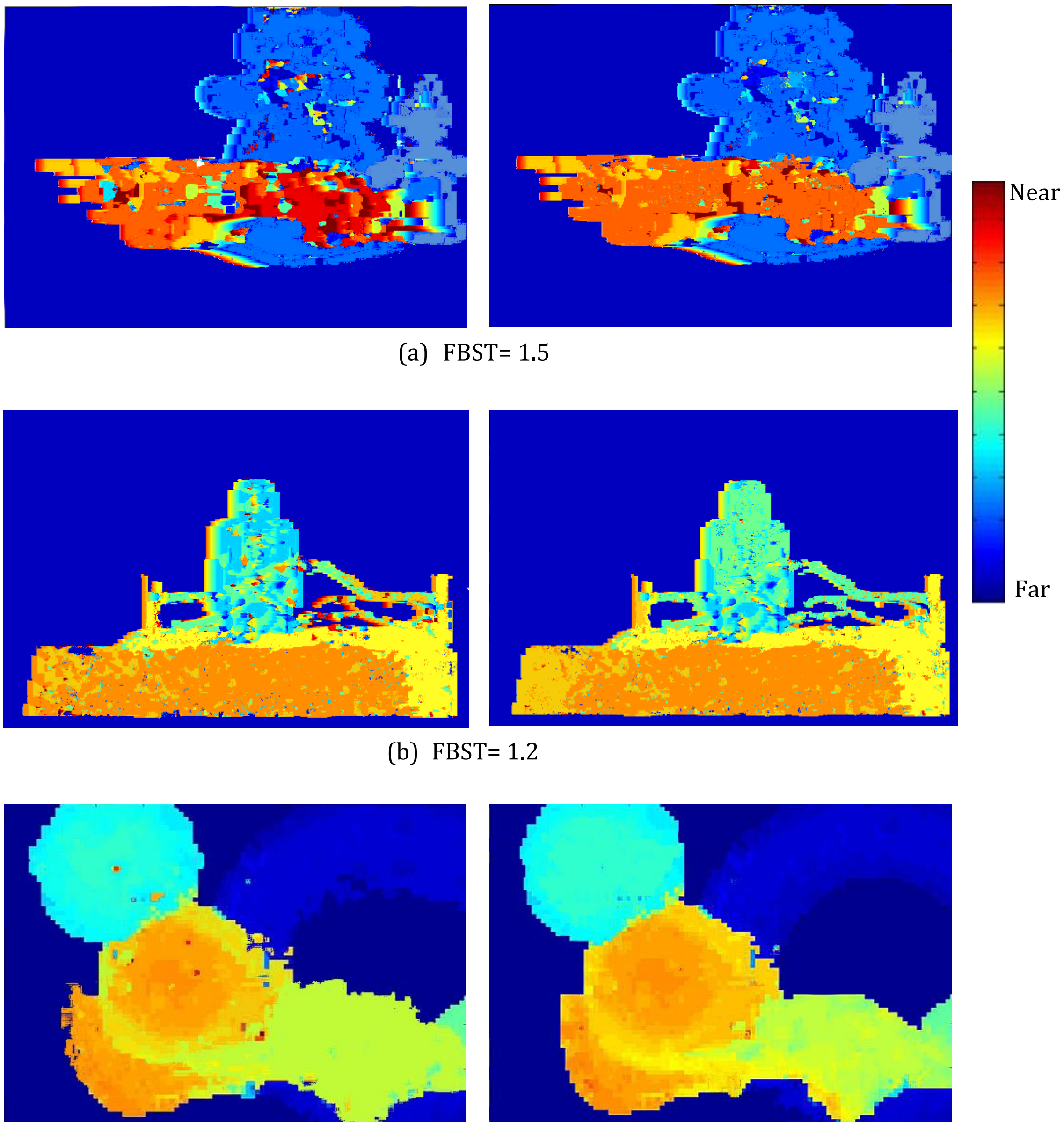


(a) FBST= 2

(b) FBST= 1.7

(c) FBST= 1.4

Figure 3-6 Comparison of the depth maps obtained from different feature block thresholds on synthetic UH3DI (bed).



(a) FBST= 1.5

(b) FBST= 1.2

(c) FBST= 1

Figure 3-7 Comparison of the depth maps obtained from *Wu et al.'s* [23-24] method in the left column and from the new adaptive weighting factors algorithm in right column. In this task the feature block threshold is preferred, a) Toy = 1.5, b) Bed= 1.2 and c) G G-Ball= 1, for a good viewing effect on the depth map.

II. Real-world UH3DIs

Encouraged by the results on the synthetic images, the algorithm has been tested on real-word UH3DIs. The "Horseman" and "Tank", both H3DIs sized 1280 × 1264 pixels, had 160 cylindrical lenses, therefore offering 160 EIs of size 8 × 1264 pixels and 8 VPIs of size 160 × 1264 pixels. The "Horseman" has some non-informative and homogeneous regions, while the "Tank" has some textureless regions (feature lacks). Figure 3.8 shows both real-world UH3DIs and one of the eight extracted VPIs for each image.

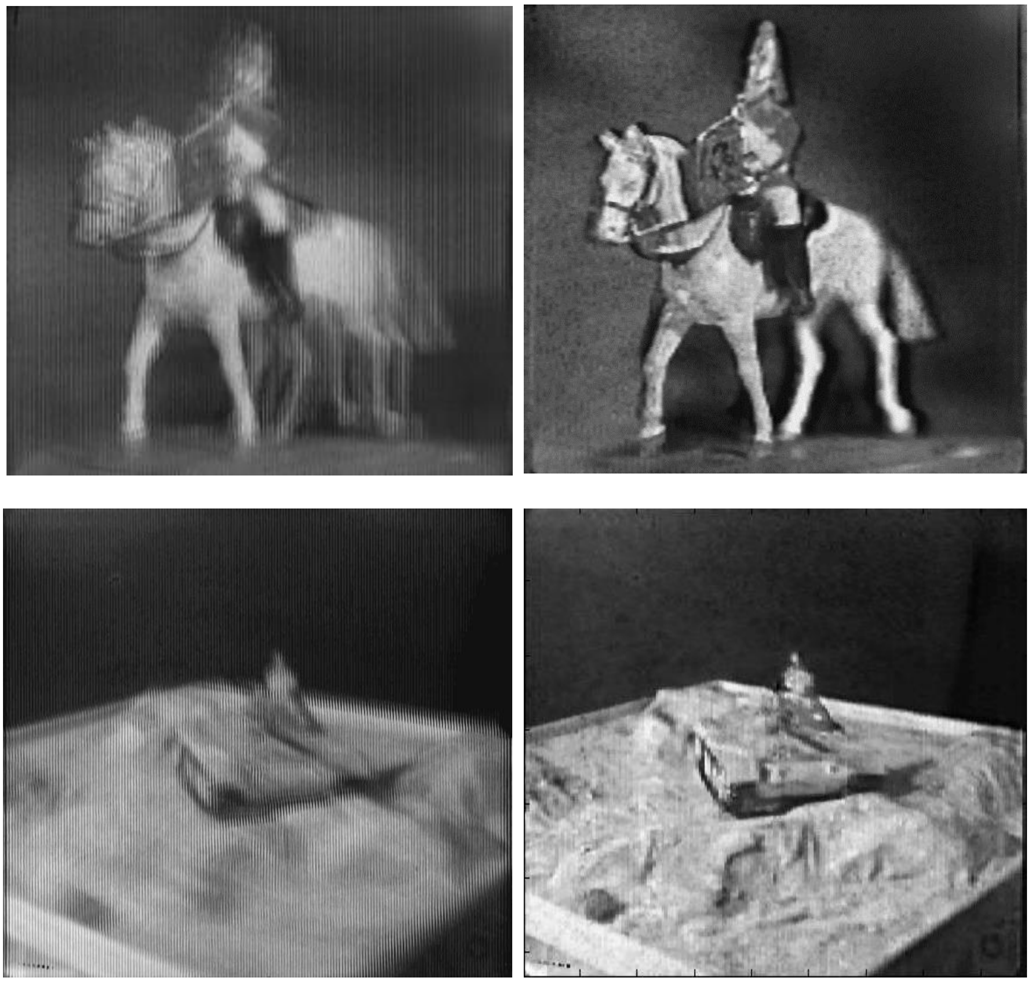
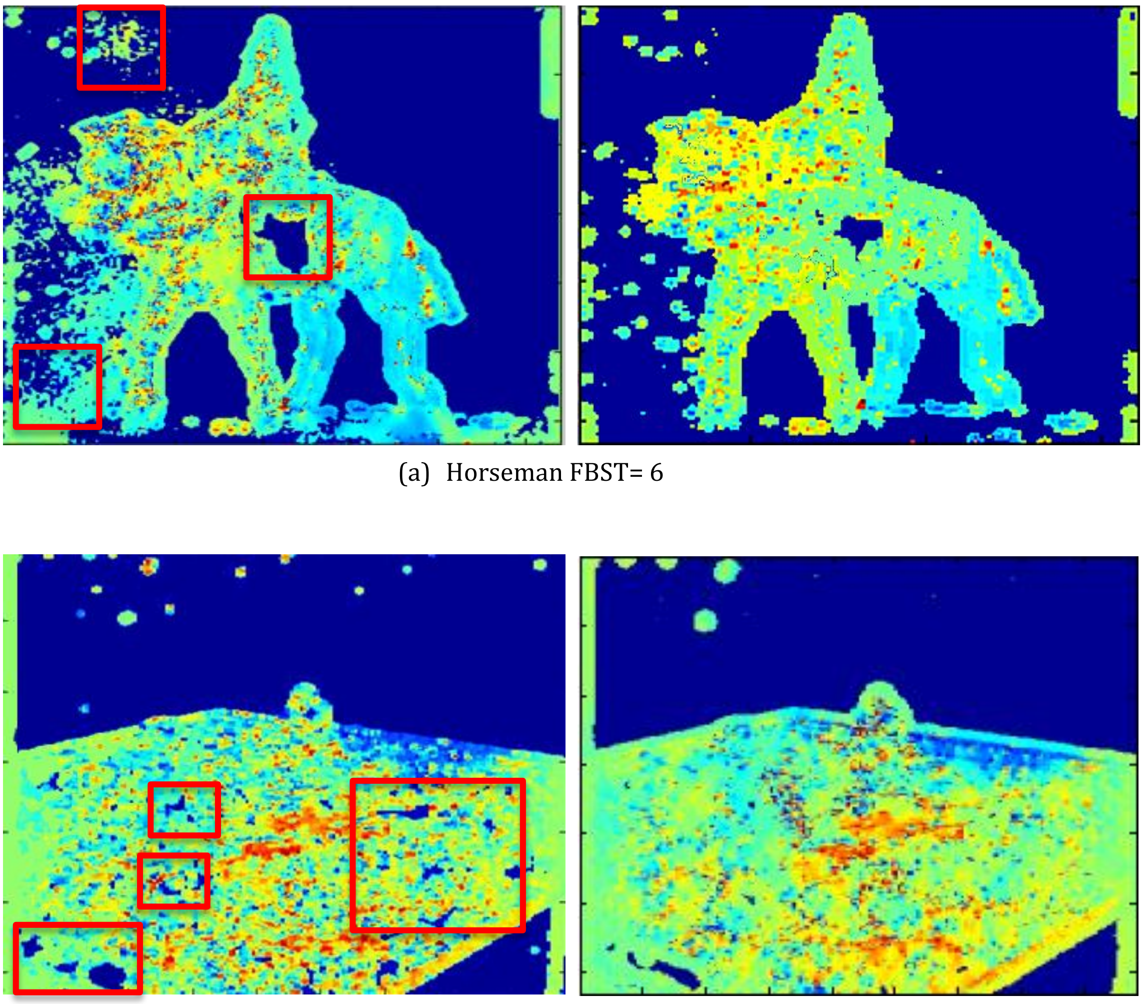


Figure 3-8 Real-world UH3DIs. The top is the "Horseman" and "Tank" is on the bottom. The right side column presents their extracted VPIs.

The final depth maps obtained from the method in [23-24] are shown in Figure 3.9 (a) using the parameters listed in Table 3.2. The result illustrates that the depth variations within objects are perceivable with acceptable quality. However, the algorithm performs poorly for regions with less texture. This results in holes in the disparity map and 3D object scene. Furthermore, the disparity map is noisy along the edges. There are potential ambiguities in matching resulting from object surfaces at different depths that alters the contour 3D object boundary. This is because selected candidate feature blocks have certain features in the image that produce more reliable matches than others. Since it accumulates multiple VPIs, sufficient precision is still achievable while avoiding matching ambiguities. Figure 3.9(a) illustrates the estimated "horseman" depth map where the head of the horse is estimated to be near the camera in yellow colour, the man light green while the tail far with blue colour. Poor results are obtained on the background region due to the lack of enough features in the region for matching. The second image "Tank" (Figure 3.9(b)) contains a tank model on a simulated sandbank basement. The variations within the tank and basement are perceivable within the colour range of the estimated depth map. The algorithm requires the presence of a detectable texture within each correlation window, therefore it tends to fail in featureless or repetitive texture environments as shown in Figure 3.9 (red squares). Moreover, the algorithm constraints are insufficient to remove the matching ambiguity for all features. Therefore, the selection of a suitable support window for aggregation matching costs ensures reliable estimation. Figure 3.9(b) shows the *desired* result, thereby reducing potential matching ambiguities of those neighbours at occlusion boundaries and within featureless regions. Experiments prove that the algorithm works effectively in achieving a good balance of reduced mismatching while maintaining a good object contour.

Table 3-2: Main parameters used to estimate final depth map from real-world UH3DI.

Main Parameters	Horseman	Tank
Number of VPI	8	8
Block size	4	5
Research area (R)	4	4
Neighbouring blocks number (N)	8	4
Neighbouring weighting factor	0.8	0.8
Extracting features blocks threshold	6	4



(a) Horseman FBST= 6

(b) Tank FBST= 4

Figure 3-9 Comparison of the depth maps obtained from *Wu et al*. [23, 24] method in the left column and from the new adaptive weighting factors algorithm in right column. In this task the feature block threshold was preferred, a) Horseman = 6 and b) Tank = 4 for a good viewing effect on the depth map.

3.5 Depth Estimation from OH3DI

The principal of the OH3DI system involves the two processes of capture and display as seen in Figure 3.10. The distribution of the light rays from the object in both processes requires a square based spherical micro-lens structure, which offers parallax from all directions, i.e. horizontal and vertical direction parallax. The planar detector surface records the H3DI as a 2D distribution of intensities that are sampled in the form of an EI array.

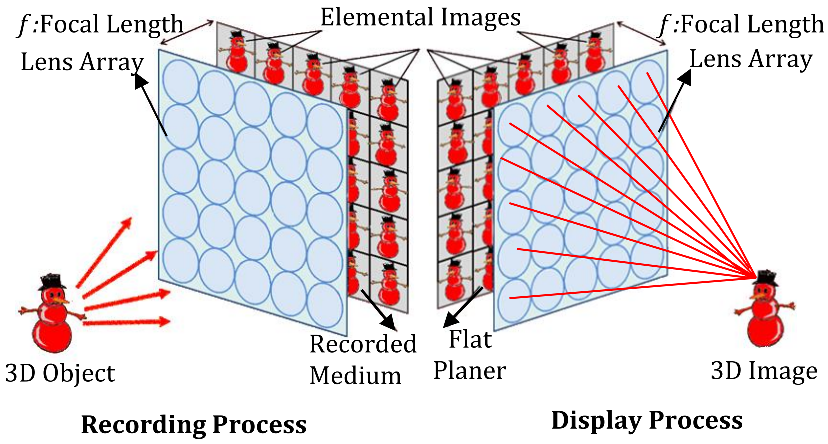


Figure 3-10 Omni-directional H3DI system.

3.5.1 Pre-processing Recorded OH3DI

Holoscopic 3D images were captured and stored in two formats (jpg and CR2) using a canon 5D camera with 90 mm lens. The OH3DI resolution was 5616 × 3744 pixels with 193 × 129 micro-lenses. This gave an EI resolution of 29 × 29 pixels while the number of micro-lenses used in the recording was used to determine the VPI resolution. Thus the VPI resolution was 193 × 129 pixels, the same as the number of micro-lenses (see Figure 3.11).

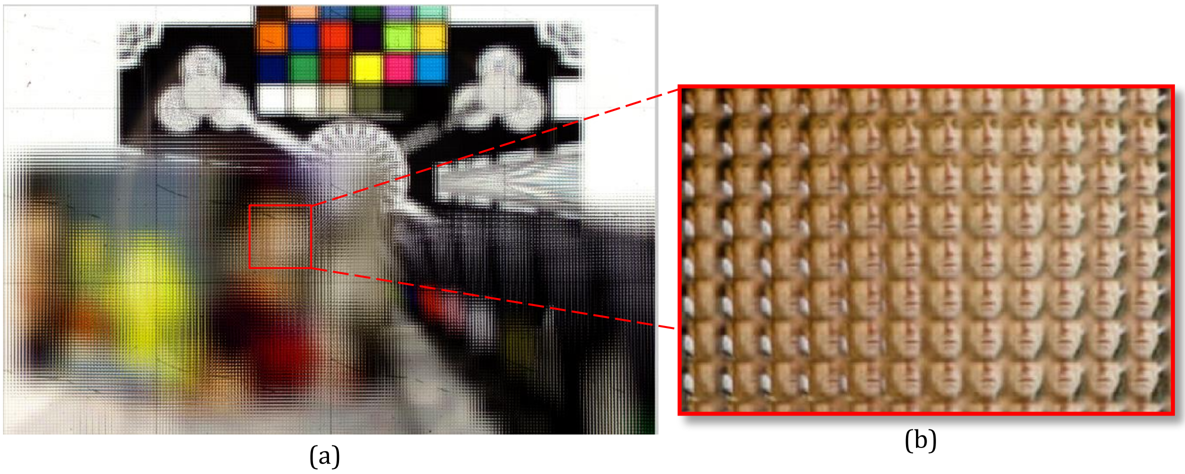


Figure 3-11 Shows real-world OH3DI captured by canon 5D camera , a) “Box-Tags” image with size 5616 × 3744 pixels, and b) magnified part of image.

Two types of distortion are commonly caused during the capture process by the lens, which are:

1. Barrel distortion: This is due to the imaging being fitted into a space that is too small. The distortion varies radially with the most obvious imperfections at the corners and sides of the images, which is due to the design of the lenses (Figure 3.12(a)).
2. Scaling errors distortion: This is when a "dark moiré effect" causes dark borders in the recorded image and errors in the playback of the 3D display. This causes errors when the 3D scene is constructed in space (Figure 3.12 (b)).

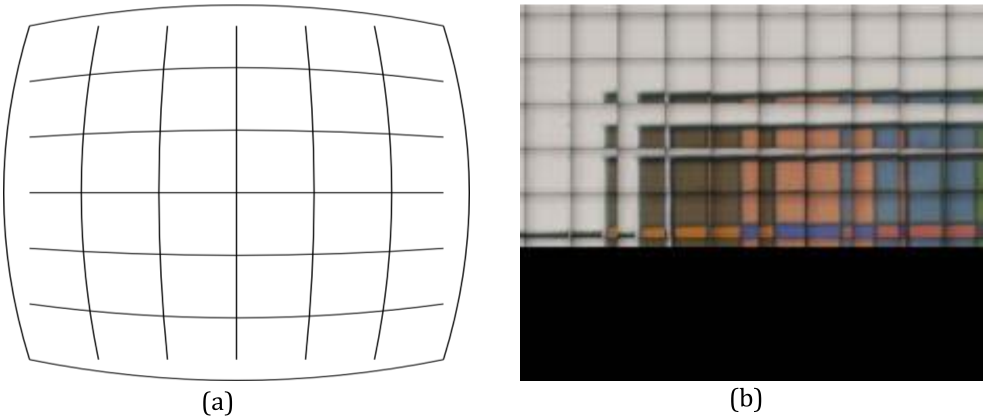
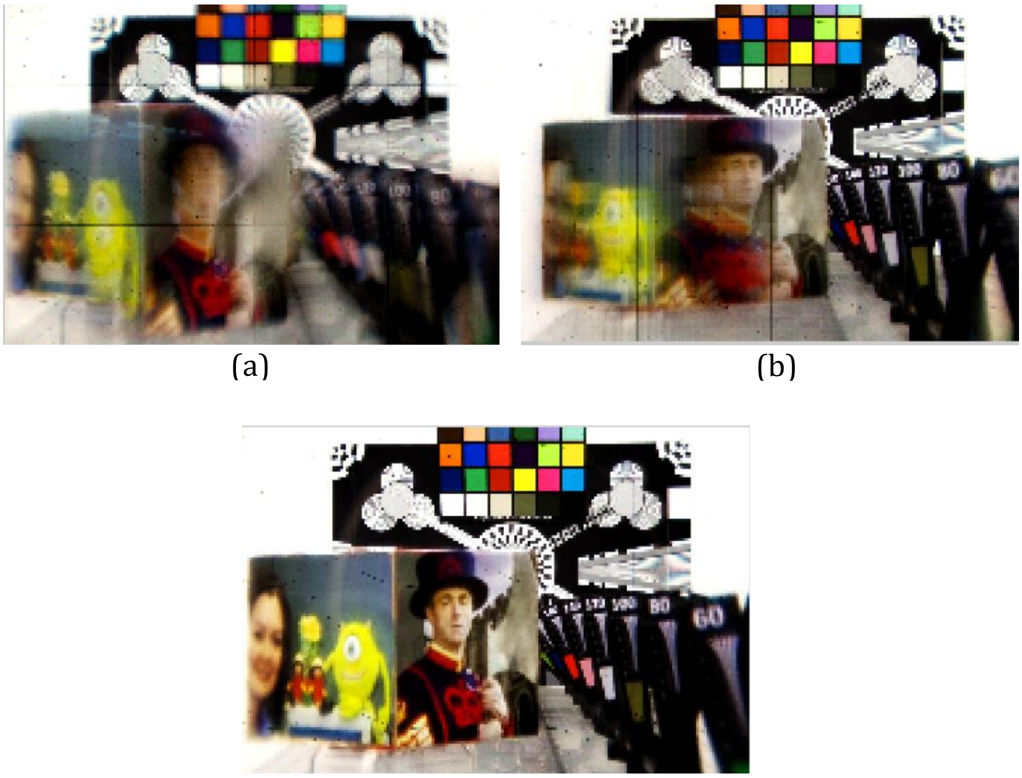


Figure 3-12 Common H3DIs distortions, a) barrel distortion effect, and b) dark borders on captured H3DI effect.

In most image applications the distortions cannot be seen by the naked eye, therefore they can be ignored. However, when extracting VPIs each pixel has to be extracted from the EIs. Thus, image distortion would cause a problem and so needs to be prevented. Examples of the VPIs extracted when the distortions have not been corrected are shown in Figures 3.13(a, b). These have unusual forms as the same pixel could not be extracted from multiple EIs.

The barrel distortion is removed with a Photoshop application. The raw data from the camera (the CR2 image) was opened in a lens correction window before it was used in the Photoshop window. In this form the data is at its most detailed as there is no compression or interpolation. If this is not ingested like this the image may be harder to correct as the intensity value for each pixel will be spread-out to neighbouring pixels, thereby decreasing the richness. The parameters can be adjusted with greater flexibility if the raw data is used. (See Appendix (A) for more details related to the correction of the barrier distortion.)

A simple and fast visual data processing process is used to correct the scaling error distortion as once the camera is setup, a simple histogram filtering algorithm generates the micro-lens grid from a H3DI of a white background and then captures the H3DI, the proposed method uses the calibrated micro-lens grid information to eliminate the dark borders of the 3D image. More details and explanation of the scaling error distortion correction is available in Appendix (B). Figure 3.13(c) shows the effects of both distortions on the VPI position (10, 10) and the accurate VPI.


(a)

(b)

(c)

Figure 3-13 Distortion effect on VPI (10, 10), a) barrel distortion effect, b) scaling error distortion effect and c) the correction VPI.

3.5.2 Transformation of OH3DEIs into Viewpoint Images

To extract VPIs from the OH3DI, the computational reconstruction algorithm generates VPIs independently by superimposing the pixels from all EIs, as shown in Figure 3.14. The OH3DI is defined as: OH3DI $= [OH3DI\ (m,n)]_{\substack{m=1,2,\dots,M\\ n=1,2,\dots,N}}$, where $m = 1, 2, \dots, M$ and $n = 1, 2, \dots, N$, and they are the horizontal and vertical positions of the OH3DI pixels respectively.

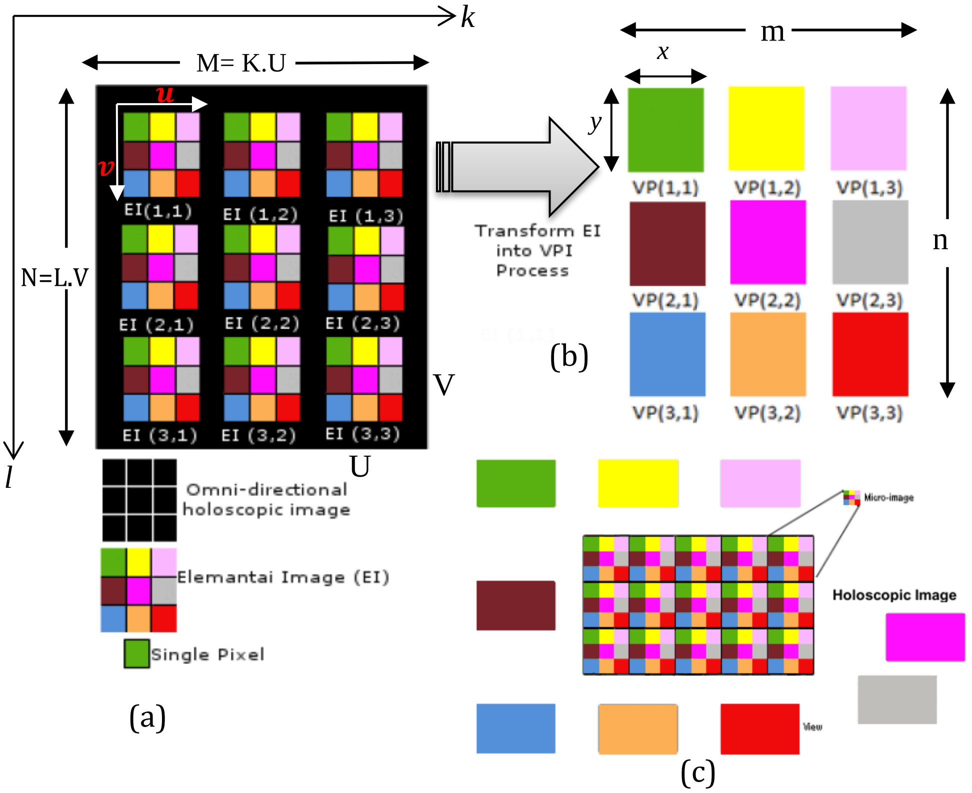


Figure 3 14 Principle of extraction for nine VPIs from OH3DI by periodically extracting pixels from the captured EI ((a-b) for simplicity, assume there are only 3 × 3 pixels under each micro-lens from 3 × 3 EIs). Extract one pixel from the same position under different micro-lenses and place them in an orderly fashion to form one VPI, and (c) extract nine VPIs from 5 × 3 EIs (i.e. m × n number of VPIs).

Furthermore, an EI is the recording image under the recording micro-lens, the OH3DI uses quadratic EIs that are uniformly positioned, with each EI at row k and column *l,* and defined as:

$$EI = [EI_{k,l}(u,v)]_{\substack{k=1,2,\ldots\ldots,K\\ l=1,2,\ldots\ldots,L}} = OH3DI\ [k.U+u, l.V+v]_{\substack{k=1,2,\ldots\ldots,K\\ l=1,2,\ldots\ldots,L}} \tag{3.17}$$

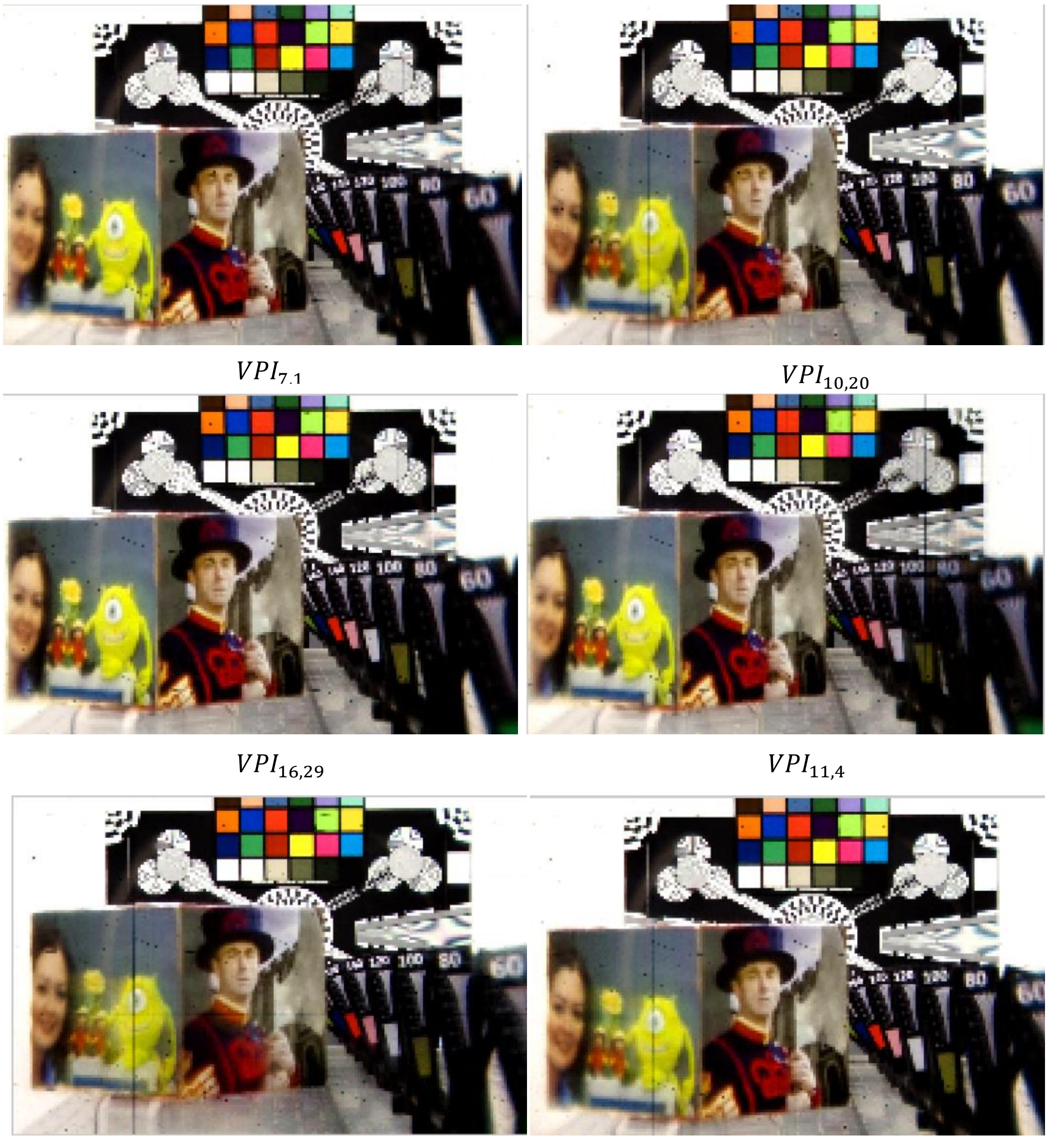
where, $u = 1,2,\ldots.,U$ and $v = 1,2,\ldots.,V$ are the horizontal and vertical pixel positions within the EI. Therefore, each $K \times L$ EI has a resolution of $U \times V$ pixels and has its *u, v* coordinate system. The VPIs are formed as subsets of pixels in the OH3DI sharing the same relative horizontal and vertical offset to each EI centre, and thus the VPI is defined as:

$$VPI_{u,v}(k,l) = [EI(u,v)]_{\substack{k=,1,\ldots\ldots,K\\ i=,1,\ldots\ldots,L}} = OH3DI\ (k.U+u, l.V+v) \tag{3.18}$$

Let the total number of EIs be $K \times L$ and a VPI (*x, y*) is evaluated by the summation of $EI_{k,l}$, which is the intensity of the *k-th* column and the *l-th* row of EL. This gives:

$$VPI(x,y) = VPI_{u,v}(k,l) = \sum_{k=1}^{K-1}\sum_{l=1}^{L-1} EI_{k,l}(x-kS, y-lS) \tag{3.19}$$

where, x and y are the index of pixels in the x and y directions, and S is an integer number of shifted pixels (pixel by pixel) in the overlapping EIs: the minimum step between the shifting distances becomes one pixel. Figure 3.15 shows the display of some extracted VPIs of the real-world OH3DI "Box".



$VPI_{7.1}$ $VPI_{10,20}$

$VPI_{16,29}$ $VPI_{11,4}$

$VPI_{18,19}$ $VPI_{10,10}$

Figure 3-14 Various display positions of VPIs extracted from the OH3DI "Box" by periodically extracting pixels from the captured EIs. It is a real-world image captured at the 3D VIVANT project lab. The resolution of the image is 5616 × 3744 pixels. A 29 × 29 grid of VPIs was extracted from the image each with 193 × 129 pixels.

3.5.3 Experimental Results

One of the most challenging tasks for the adaptive algorithm is when it is utilized on an OH3DI system to improve the ability and efficiency of the algorithm to accurately measure the 3D depth map. Most state-of-the-art algorithms for computational depth generation from H3DI systems employ their algorithms with a UH3DI system. This experiment performed on OH3DIs was based on the proven success after implementing the adaptive multi-baseline algorithm on both synthetic and real-word data UH3DIs. To the best of the author's knowledge, this process of depth estimation through computational simulation of OH3DIs has not been reported in the literature previously. Depth estimation from H3DI system in most of the literature has been addressed as an optical hardware issue. In this section the experiments were carried out on two types of OH3DIs: computer generated and real-world images.

I. Synthetic OH3DIs Database

The approach outlined above was used to test the feasibility of depth estimation from computer generation OH3DIs. A "3DVIVANT" OH3DI that contained the object scene was computer generated using a software tool developed by the Brunel team, which involves a 3D holoscopic virtual camera containing a micro-lens array with a 150 × 150 micro-lens. Figure 3.16 shows the computer generated OH3DI and a magnified part. The resolution on the computer generated OH3DI is 4500 × 4500 pixels, with 150 × 150 EIs with a resolution of 30 × 30 (900) pixels each, corresponding to a 1:1 ratio and therefore each VPI = 150 × 150 pixels. The micro-lens pitch size was (P) = 90 mm, focal length (f) = 0.001 mm and pixel pitch (ρ) = 3.1 mm.

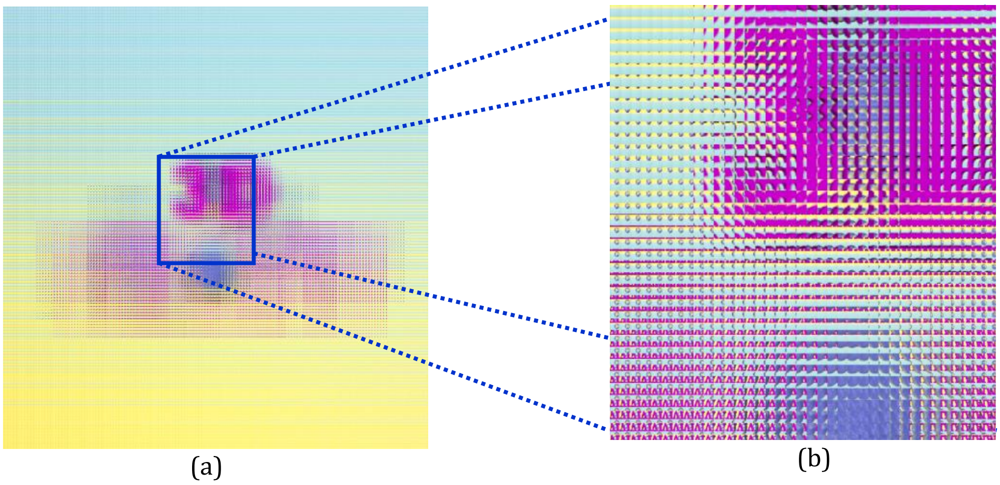


Figure 3-15 Display a) CG-3DVIVANT image and b) magnified part.

Figure 3.17 shows one of the 900 views that can be extracted from the "3DVIVANT" data, the objects are placed at an accurately measured distance from the camera. The "4

Globes", "3D" and "VIVANT" objects are placed in front of a plane background. The background plane has some pattern colours to facilitate the disparity analysis. The recorded OH3DI therefore has seven depth planes. The distance for each object, also referred to as a 'Target', is recorded from the camera's micro-lens as listed in Table 3.3.

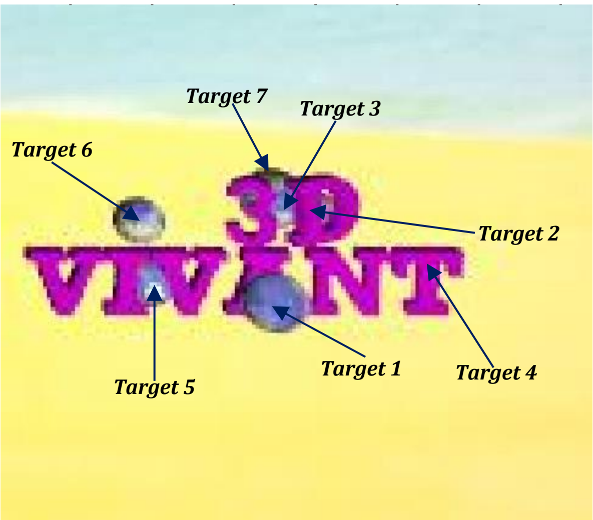


Figure 3-16 One of the 900 VPIs extracted from the CG-3D VIVANT image with a resolution 150 × 150 pixels. The targets are numbered based on the distance from the camera micro-lens.

After the transformation of all the pixels located in the EI array into the VPI array, a block array with 7 rows by 7 columns picked up from the 3D object located at the centre axis of the pick-up system of 150 by 150 arrays of VPIs is produced. It is found that the object size is invariant even though the object location relative to the micro-lens array has shifted, and the 3D object imagined on the VPI array is shifted as the 3D object is moved on the pickup condition [134]. Therefore, a highly focused image containing most of the perspective data from a 3D object at the centre of the VPI array and focused VPIs can be used for a reference VPI to be compared with the other rearranged VPIs.

Applying the adaptive weighting factors obtains the accrued score functions from the adapted multi-baseline algorithm using Eq. (3.16). The parameters listed in Table 3.4 were used to estimate the final depth map, and were considered and confirmed as being suitable parameters experimentally. The estimated depth results of the CG-3DVIVANT targets are listed in Table 3.4. The Mean Relative Error (MRE) was used to evaluate the efficiency of the algorithm and has been defined as [135]:

$$\mathrm{MRE} = \frac{100\%}{\mathrm{N}} \sum_{(x,y)}^{N=T_{depth}(x,y)} \left| T_{depth}(x,y) - E_{depth}(x,y) \right| \qquad (3.20)$$

where, $T_{depth}(x, y)$ and $E_{depth}(x, y)$ are the actual depth and estimated depth at the (x, y) pixel position, respectively, and N is the amount of compared pixels. Figure 3.18 shows the depth maps obtained from the computer generated OH3DI data. The result illustrates that the depth variations within objects are perceivable with acceptable quality. However, the algorithm performs poorly for regions with occluded objects. These results identify this as the yellow rectangles in the disparity map and 3D object scene. Furthermore, the disparity map is noisy along the edges. There are potential ambiguities in matching results (red rectangles) from objects at different depths that alters the contour 3D object boundary. This is because of selected candidate feature blocks in the matching and corresponding process.

The mean relative error (MRE) increases as the object's distance from the camera also increases [136]. Generally, this error increases parabolically with the distance, whilst it remains low for distances below 100 mm. Experimental results have shown notably, that the use of the proposed AWF algorithm has significantly reduced the percentages of MRE (see Table 3.4). Therefore, a significant improvement has been shown on the quality of output depth estimation maps (see Figure 3.7). It should be noted that there exists a dependency between the depth of an object and its corresponding resolution in the image i.e. far and near objects are seen as small and large in the image, respectively. It is apparent that the accuracy of the method is directly related to the actual object's distance from the camera. More specifically, the closer the object is to the camera the better the estimation. This is again because of the inherent ambiguities of the setting feature blocks.

Table 3-3: Depth measurement results on CG-3DVIVANT OH3DI.

Target	Max Distance (mm)	Min Distance (mm)	Actual Depth	Estimated Depth	MRE(%) before AWF	MRE(%) after AWF
1	80	48	32	29.4	8.12	5.32
2	90	66	24	22	8.33	5.91
3	100	78	22	20.1	8.63	6.57
4	120	94	26	23	11.5	7.35
5	135	102	28	23.8	15	9.58
6	168	139	29	23.68	18.34	11.23
7	186	153	33	26.35	20.14	12.08

Table 3-4: Main parameters used to estimate the final depth map from OH3DI.

Main Parameters	3DVIVANT	Airplane-man	Box
Number of VPI	49	49	49
Block size	11	11	11
Search area (R)	3	3	3
Neighbouring blocks number (N)	8	8	8
Extracting features blocks threshold	5	4	1.5

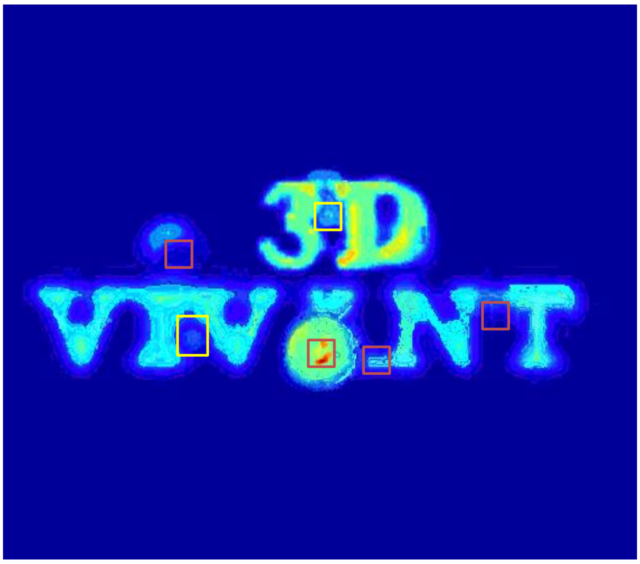


Figure 3-17 Depth results on CG-3DVIVANT OH3DI.

II. Real-world UH3DIs

To evaluate the generalization and the efficiency of the approach, the algorithm was implemented with real-world OH3DIs. Two OH3DIs "Box" (see Figures 3.11 and 3.15) and "Airplane-man" (see Figure 3.19) were captured by the "3D VIVANT PROJECT" using a 5D canon camera with 50 mm main lens and 21-megapixel size image. The resolution of each image was obtained at 5616 × 3744 pixels, which consisted of 193 × 129 micro-lenses, with an EI resolution of 29 × 29 pixels and VPI resolution determined by the number of micro-lens contained in the recording, thus the VPI resolution was the same as the number of micro-lenses i.e. 193 × 129 pixels. The parameters, which were empirically determined and not specifically adapted to each data set, are listed in Table 3.4.

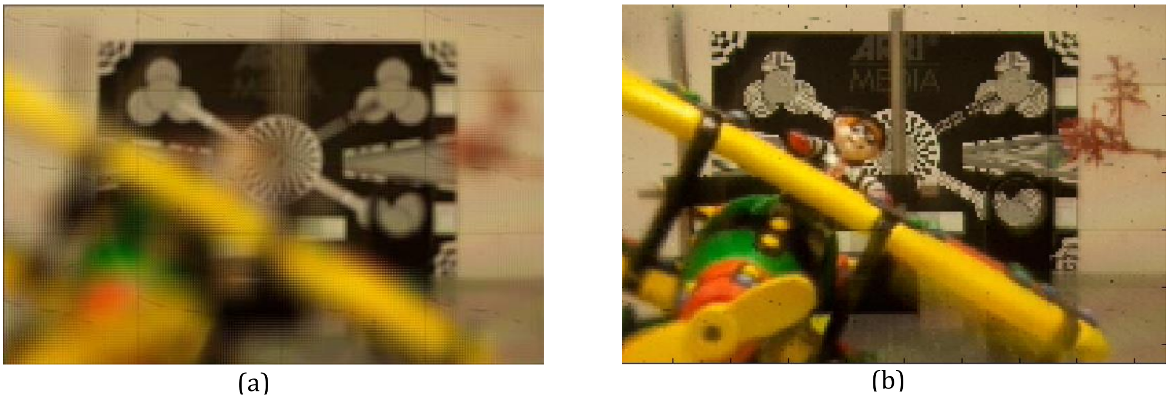


Figure 3-18 Real-world OH3DI captured by canon 5D camera , a) "Airplane-man" at 5616 × 3744 pixels resolution and b) one extracted VPI of a 841 low resolution orthographic VPI at 193 × 129 pixels resolution.

Figure 3.20(a) shows the estimated depth of the "Airplane-man" images without an adaptive shape weighting window. This results in holes in the disparity map and 3D object scene within the object near the camera because of inconsistent matches, which creates holes in the region, thus in the result no disparity values are assigned in the disparity map. While the background is occupied with false matching due to the inherent ambiguity in matching, this is further complicated by phenomena such as occlusion and untextured regions. Better depth estimation quality results can be achieved by employing an adaptive shape weighting window as shown in Figure 3.20(b).

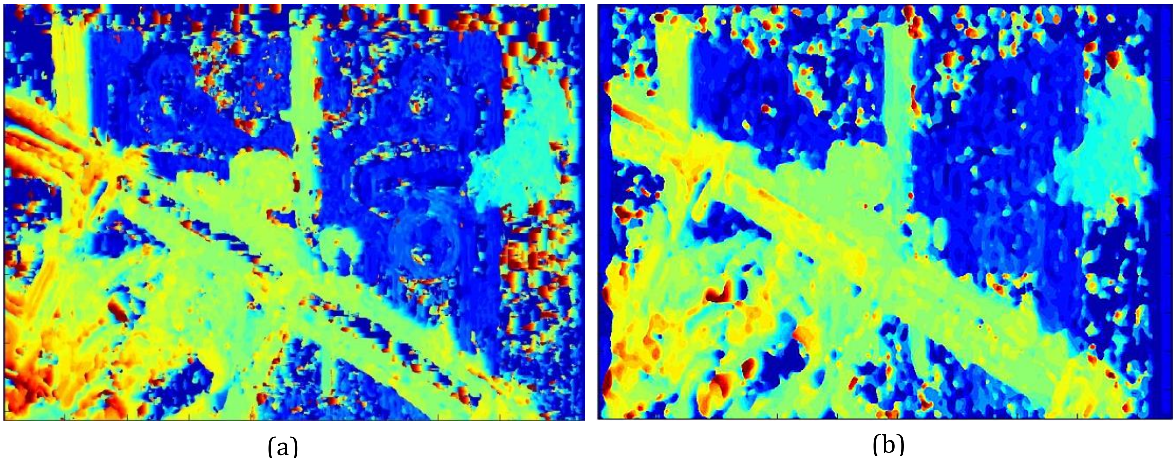


Figure 3-19 Real-world OH3DI "Airplane-man" depth map results, a) depth before employing adaptive shape weighting window and b) depth after employing adaptive shape weighting window.

Further tests were carried out on the "Box-Tags" OH3DI with different targets (e.g. variations in size and textures) having various depth ranges. Tags starting at 600 mm from the camera to a background distance of 3000 mm from the camera with 200 mm steps between each tag.

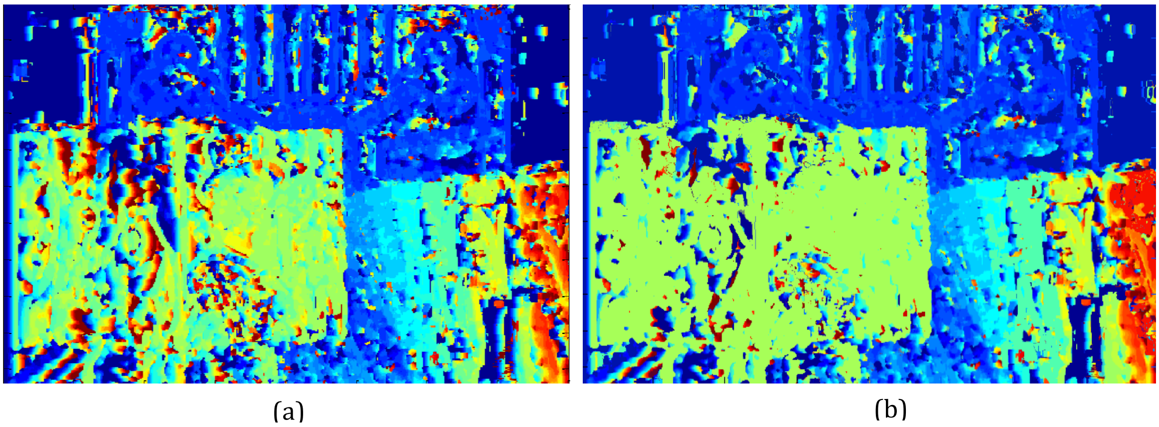


Figure 3-20 Real-world OH3DI "Box" depth map results, a) depth before employing adaptive shape weighting window and b) depth after employing adaptive shape weighting window.

In addition, there was a different texture OH3DI "Box" with 1200 mm distance from the camera. The visual quality of the resulting 3D depth maps are shown in Figure 3.21. The adaptive algorithm, using a multi-baseline disparity with adaptive shape weighting window, was subjectively tested against the original depth map based on *Wu et al.* [23, 24]. The accuracy performance of the depth map improved, though there was some sacrifice of the neatness in the background with some holes. However, increasing or decreasing the setting of and extracting the feature block threshold results in holes and ambiguities, feature blocks at different depths start to enter the matching window and make correct matching difficult to detect, thus invalid feature blocks cannot be properly recovered from the neighbouring blocks.

Another important aspect is the fact that, both the size and the texture of the sought object play an essential role in the calculation of its spatial information, in addition to the distance from the captured camera. For example, location assignments for the target "Box", which is the largest object in the scene with the most texture, are more effective due to the fact that the "Box" attributes provide more features to extract. As a result, the greater the number of features the better their distribution over the surface of the object. Since the algorithm is mainly based on the feature block spread, the enhanced distribution of the setting of reliable features can lead to more qualitative and quantitative results.

3.6 Summary

This chapter presents an approach that improves the quality of the visual depth maps in UH3D and OH3D images, the two types of H3DIs. Using a set of orthographical projection low resolution VPIs a new modified multi-baseline algorithm that contained an adaptive weighting factor shape window technique was used to enhance the estimation of the 3D object's position. First, the correlation between the feature's information points in several VPIs was used to modify the multi-baseline algorithm. To obtain depth from 2DVPIs, the first step is to select the features, before optimizing the corresponding and matching process to produce the 3D depth map. Multiple neighbouring blocks were prevented from being identified as interesting in the same feature through the use of a novel adaptive weighting factor algorithm that introduced refinement to the disparity map. In addition, this improved the computation efficiency. An adaptive aggregation cost window was used to identify large support areas in untextured zones as well as adapting shapes that appeared near the boundaries of objects. This was achieved by ensuring that extra emphasis was placed on the center block rather than surrounding blocks. This was tested via computational experiments that were performed using both computer generated and real-world UH3DIs. This gave a successful result in which both object recognition and position estimation were improved.

In the chapter's second section, the computational simulation of depth estimation was studied on both computer generated and real-world-images in which the estimation of 3D depth from OH3DI is considered. This had not been previously covered in the literature as most work was related to the estimation of 3D depth on UH3DIs. The capture of OH3DIs introduced geometric distortions that had to be corrected before it was possible to transform the EIs into VPIs. This had to be done to ensure that it was possible to extract the same pixel from multiple EIs. The most common argument against the use of the multi-baseline approach is, "occlusions become an issue for large baseline" [126]. This problem was overcome by using a multi-baseline technique where the center VPI was used as a reference image from which the surrounding VPIs were selected as target images. This new method was evaluated experimentally by using complex object scenes that were computer generated or real-world OH3DIs. The method proposed here was an obvious improvement over the method proposed by *Wu* *et al.* [23, 24], which was due to it incorporating an adaptive weighting factor shape window and multi-baseline technique.

The proposed algorithm was proven to be effective via experiments on both UH3DIs and OH3DIs. The approach forced the algorithm to work with complex object scenes in OH3DIs by overcoming the problem of generalization capability. In addition, the generated depth maps have improved the performance. However, there are limitations and these come from the loss of information due to the setting and extracting of reliable features, with the most noticeable errors occurring in the background region as there is a lack of features to make the match. The threshold for the setting of the features is determined manually for each image, which is very time consuming, as there are differences between all the images that could result in false matching and faults in the depth estimation process.

The disparity map's performance is an essential part of obtaining accurate depth estimation. This work developed and implemented a new corresponding and matching technique that was based on automatic 3D feature detectors and descriptors in the thresholding that improved the performance (generalization and quality) of the depth estimation.

CHAPTER 4

Hybrid 3D Depth Estimation Techniques

4.1 Introduction

In general, manual, semi-automatic and automatic are the three categories that depth map estimation techniques can be classified into. In the manual method, the outline of the chosen object would be drawn by an operator. When the depth map is estimated with the through-disparity-analysis algorithm the first step is often performed manually as this is the hardest. Chapter three used the manual initialiser method, and the depth map is estimated from the amount of feature descriptors used. This system is based on databases that contain many more feature descriptors than are needed. This means that the methods are time consuming and expensive, and therefore not suitable for real-time applications. This affects the quality of the 3D models in the reconstruction process as it takes a very long time to search a very large database. This could be solved by removing the least important feature descriptors by selecting distinguished feature descriptors. This means that when estimating the depth map either a semi-automatic or automatic technique is preferred. The human visual depth perception mechanism is used as the base from which these techniques are designed.

In this thesis, the main topic is to develop automatic initialization techniques for creating and manipulating 3D models from a H3DI system that are fast and accurate. This will enable specific points in the reference VPI to be found and then matched to other targets in an image sequence via reconstruction algorithms. These features are then created in 3D by the H3DI display system so the viewers can see and appreciate an object or scene fully.

The features in a 2D image have to be known before they can be used to undertake depth estimation. For this a feature that has the characteristics of being repeatable and distinctive, is a good feature. For it to be repeatable it must be possible to extract it from multiple representations of the same scene. A distinctive feature means that it will closely correspond to a matching feature in another image. The information on these characteristics has to be stored in a way that is unique to that feature. In addition, the descriptor should not be influenced by the position of the camera. For a feature to be identified as known, its descriptor has to match the known feature via the matching process.

This chapter proposes two novel automatic local feature setting and selecting techniques that are based on binary data. These will often be edges or points of the 3D object. Two techniques (auto-feature-point setting and auto-feature-edge detection), which identify the known features through a learning methodology, are reviewed in section 4.2 and 4.3, respectively. Once suitable matching features are identified, the adaptive multi-baseline disparity algorithm (see Chapter 3) is implemented to determine the matching features in multi pairs of VPIs by comparing the relation between the features.

The aim when detecting and then setting 2D features from the input VPIs is to obtain a set of matching pixel positions across all the images. The depth map is then developed by converting the 2D features into a 3D space, but this is influenced by the content of the images. This depends on the scenes' complexity. Corresponding features to be matched are required them to be unique. This means that the content must have descriptors that are distinctive and easily recognisable throughout the set of VPIs. This limits the content that can be used with 3D reconstruction.

4.2 Feature-Point Setting Technique

Points of interest on the 3D object are the local feature points that can be extracted and matched between VPIs reliably, and they are either a point (Interest Point) or group of pixels (Region of Interest). The local feature selection, description and matching are regarded as low level processes that follow a bottom-up approach. Once they have been completed the higher level processes are conducted where this basic information is converted until a conclusion at a semantic level is reached. The meaning is recovered, recognition performed or a high-level structure is estimated at this stage. The optimal current descriptor approaches all use binary string descriptors where the vector dimensions of the descriptor are recorded as a string of 1's and 0's [137-140]. The use of these descriptors is less demanding of computer processing than when real value descriptors are used, which is due to the ability to compute the Hamming distance between two binary descriptors [140, 141].

Feature extraction is the first and most important step in depth map estimation when the extracted image features from the 2DVPIs are matched. Rather than use a discrete or sparse set of points, the relative depth is estimated for each pixel in the VPI. This can only be used for short base-line applications where the difference between VPIs to be compared is small. In a wide baseline image it is not possible to use the image patch individually as this is inefficient due to instability and sensitivity to small changes, as well as lack of robustness against geometric or photometric transformations and reduced distinctiveness capabilities. Therefore, better strategies are needed to identify, distinguish and match the points from the various possible matches.

To set and detect feature blocks from H3DI automatically a new and simple Feature-Match Selection (FMS) technique has been developed. The FMS algorithm is an efficient and informative procedure implemented via assessing the intensity variance of the image blocks before disparity analysis. A block is considered as "untraceable" if the intensity variance within the block is smaller than a given threshold. Those blocks containing sufficient features for a confident match are chosen for further analysis. The depth of the chosen blocks can be recovered from the neighbouring blocks.

The spatial intensity distribution in an image is used as features to represent an image. In order to extract the features from 2DVPIs, a threshold giving highest local contrast is required. False matching from this, leads to an incorrect depth analysis process by the untraceable block that is identified via the feature block. To ensure good efficiency in the matching, these types of blocks should be identified before starting the matching process, which can be achieved by considering the variance of the blocks. The matching results from these untraceable blocks will then be treated as low confidence and discarded, so that only blocks with confident matches are retained.

The technique is novel due to two aspects: the corresponding/matching process on the 2DVPI's for the selection and optimization of the feature block selection. Thus solving of the three main problems related to depth map estimation from H3DI system requires: i) that at the image location where errors are frequently found in the disparity process has uncertainty and region homogeneity, ii) dissimilar displacements within the matching block around object borders and iii) object segmentation.

4.2.1 Optimal Threshold Technique

The principle of the method is to search for the optimal threshold value that gives the highest local contrast by comparing small patches extracted at each region in the image to their immediate neighbourhood. The spatial intensity distribution in an image is used as features to represent an image. The threshold that gives the greatest local contrast is needed to obtain the features from a 2DVPI. The principle of this method is depicted in Figure 4.1.

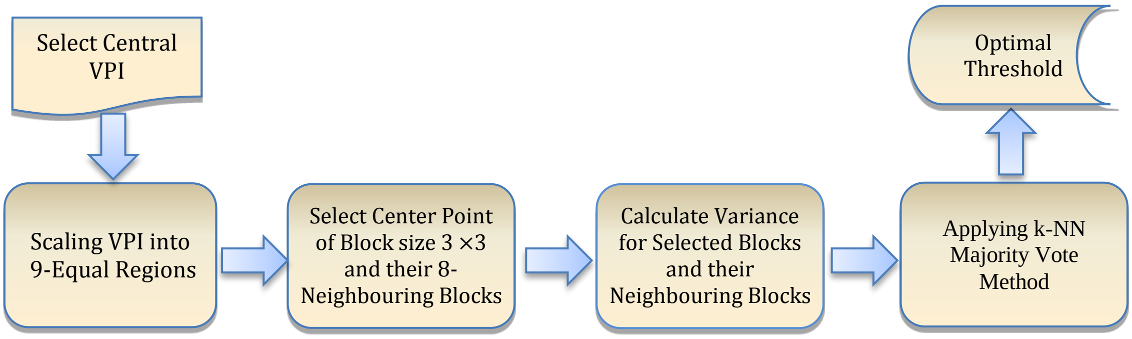

Select Central VPI
Scaling VPI into 9-Equal Regions
Select Center Point of Block size 3 ×3 and their 8-Neighbouring Blocks
Calculate Variance for Selected Blocks and their Neighbouring Blocks
Applying k-NN Majority Vote Method
Optimal Threshold

Figure 4-1 Pipeline mechanism for Optimal Threshold Technique (OTT).

The optimal threshold algorithm is described in the following steps:

1. Select Reference Images

Select the centre part of the extracted VPIs as a learning image. This is due to the over or under-exposure near the edges of micro-lenses during the capture process, which means the extracted VPI would be either too bright or too dark.

2. Up-sampling Viewpoint Image

Each extracted VPI's resolution is equal to that of the micro-lens array (or sheet), which means they are low resolution images. To obtain a high-resolution information image, the training reference VPI is up-sampled by the number of input VPIs used to estimate the disparity map via fast bi-cubic interpolation [142]. The highest repeatability is obtained when up-sampling scales capable of generating sharp edges with reduced input resolution grid-related artefacts [143]. When performing up-sampling, then more pixel intensities than the given number take place. From this, the number of correct matches increases as the number of feature points increases.

3. Sampling Viewpoints Image

The VPI is sampled to speed up the FMS process by minimized inclusion of data from different classes. The up-sampled VPI is firstly divided into 3×3 non-overlapping square regions, in which different sized blocks and a grid can be used, as shown in Figure 4.2(a).

4. Select First Point and K-nearest Neighbour

The initial testing point, for each region, in the training VPI is one randomly selected individual sample block (offspring) Figure 4.2(b). The other blocks will be around the sequence nearest neighbour (NN) blocks (Figure 4.2(c)).

5. Generate Variance Vector

Then for each block (see Figure 4.2(d)) determine the intensity variance δ as its representative feature. Each block has an area of 3×3 pixels. Let I_i be the intensity of each block and B as the number of pixels in each block (e.g., nine in this case). Then the following gives each block's variance (δ):

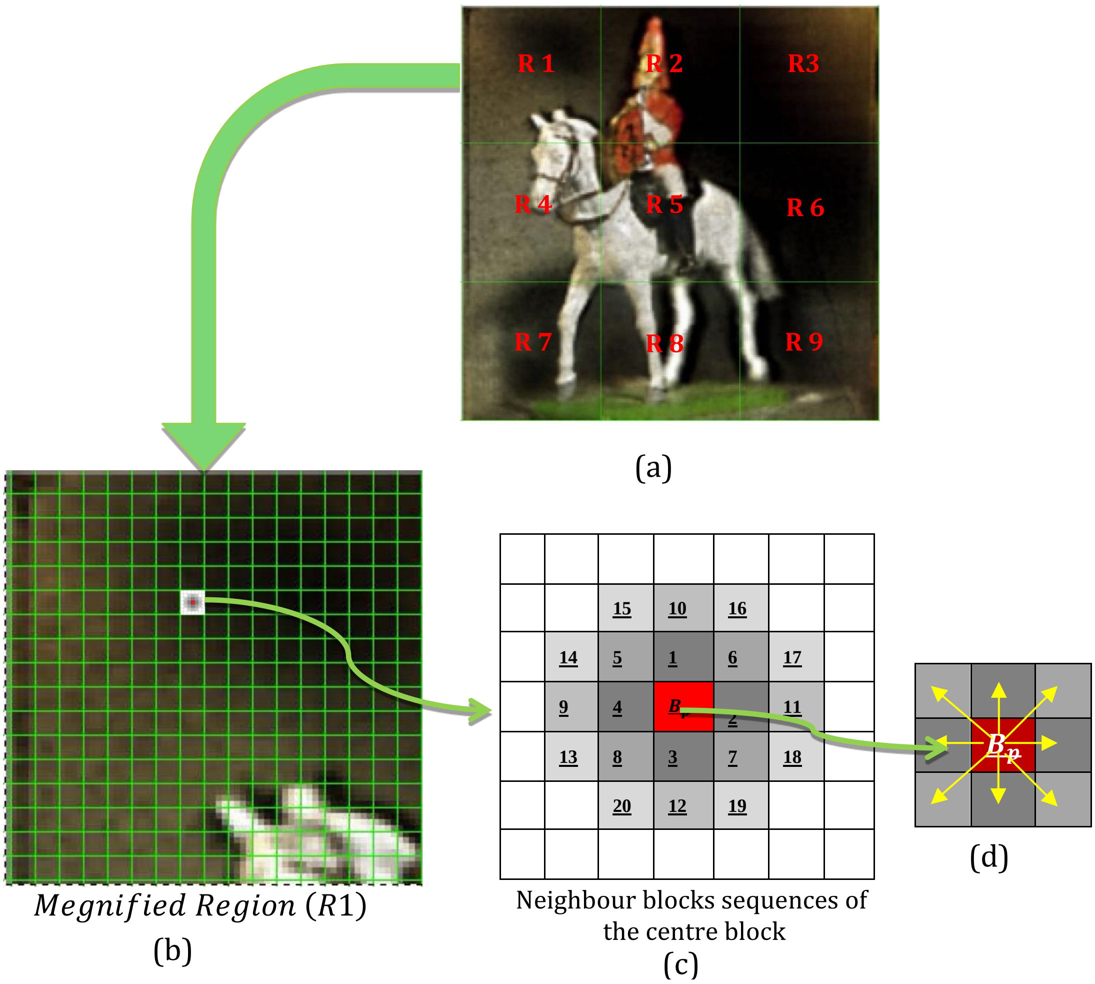
$$\delta = \frac{1}{B-1} \sum_{i=1}^{B} (I_{i-}\mu)^2 \text{ , for } i = 1,2, \ldots \ldots ,8 \qquad (4.1)$$

where each block's mean is: $\mu_i = \frac{1}{B} \sum_{i=1}^{B} I_i$

For the centre of the block and the neighbour blocks the vector of the variance is $\delta = [\delta_1, \delta_2, \ldots, \delta_9]^T$. Without loss of generality, for the centre block the variance is δ_1 and any one neighbouring block as δ_2 and so on until the last neighbour has been denote.

6. Applying *K-NN* Majority Vote Method

Of all the machine learning algorithms that have been developed one of the simplest is the k-nearest neighbours (k-NN) algorithm. Its central assumption is that class probabilities are approximately constant at least locally, but for most neighbourhoods this is not true. Therefore, in new neighbourhoods the class probabilities need to be made "more" constant via changing the metric (distance) and the majority will vote for the k closest training instances.



(a)

Megnified Region (*R*1)

(b)

Neighbour blocks sequences of the centre block

(c)

(d)

Figure 4-2 The optimal threshold Technique, a) Nine sub-divided non-overlapping regions on central sampled “Horseman” VPI, b) A sampled randomly selected block in red and *k-NN* blocks sequences on first region, c) The *k-NN* blocks distance –weight direction of the centre selected point block.

For records that have been stored in ascending order the distance between them can be computed using the distance matrix vector d in d_m. Let $\mathcal{L} = [d_m, \delta_m]$, in which the corresponding vector of variance δ_m is based on the order of d_m. Each region’s features are classified via the majority from the votes of the k nearest neighbours.

The following have been proven by experimental results in this work:

- When k is too small it will be sensitive to noise points.
- When k is too large there maybe points from other classes included in the neighbourhood.
- k should be an odd value so that ties will be eliminated

From δ_m the first three variances should be chosen, which are representative of the sample random block’s (offspring) variance.

$$\delta_{m,i} = arg\,max_k\{(\delta_{m,i})\} \;\; \forall \;\; \delta_m \; => \delta_c \tag{4.2}$$

where, δ_c is the variance of the centre block.

Voting is conducted to determine which class the majority of the k's vote for in the k-NN. In the case of a tie the smallest k is the tie breaker. The average of the majority should be calculated, as in the variance feature $\delta_{m,i}$ that has obtained the greatest amount of k-number "votes" from all of the vectors δ_m related to the sample centre point block b_p. That is, these are all the selection features that apply to the same centre block. Then, the representative feature for the first region is determined from the average of these three variances v_1.

$$v_1 = \frac{1}{k}\sum_{i=1}^{k} \delta_{m,i} \tag{4.3}$$

If the block's variance is $\delta_{m,i} = 0 \;\; or < \delta_c$, when $\delta_c = \delta_1$ is the center point's variance, there is no feature block. When this occurs the next nearest neighbour should be identified and the above procedure repeated.

7. Optimal Threshold (OT)

To calculate $v = \{v_1, v_2, \dots, v_M\}$, all the steps in the k-NN procedure should be repeated for all regions $1{\sim}M$ in the reference VPI. Then calculate the mean variance v as in the following to obtain the optimal threshold OT*:

$$OT^* = \frac{1}{M}\sum_{i=1}^{M} v_i \; , \; i = 1, 2, \dots, M \tag{4.4}$$

4.2.2 Generate Feature Block Profile (FBP) using Optimal Threshold

The procedure for setting and extracting individual feature-point profiles will be considered in this section. Binary decision descriptors in a look-up table are used for further analysis of the standard feature point's profile that represents a particular feature. Each dimension of the vector that describes this descriptor is a binary bit string i.e. "1" or "0" that depends on the distribution of the optimal threshold along the VPI's patch (block) in the form of:

$$\Upsilon\,(\beta; x, y) = \begin{cases} 1 & (\beta; x, y) > \; OT^* \\ 0 & otherwise \end{cases} \tag{4.5}$$

where, for pixel point (x, y) the intensity value is represented by $(\beta; x, y)$. The binary descriptors are less demanding computationally than real value descriptors, resulting in greater efficiency in the matching process. The method that is used to locate and then extract all the valid feature-points to form the accumulator vector from the profile is shown in Figure 4.3.

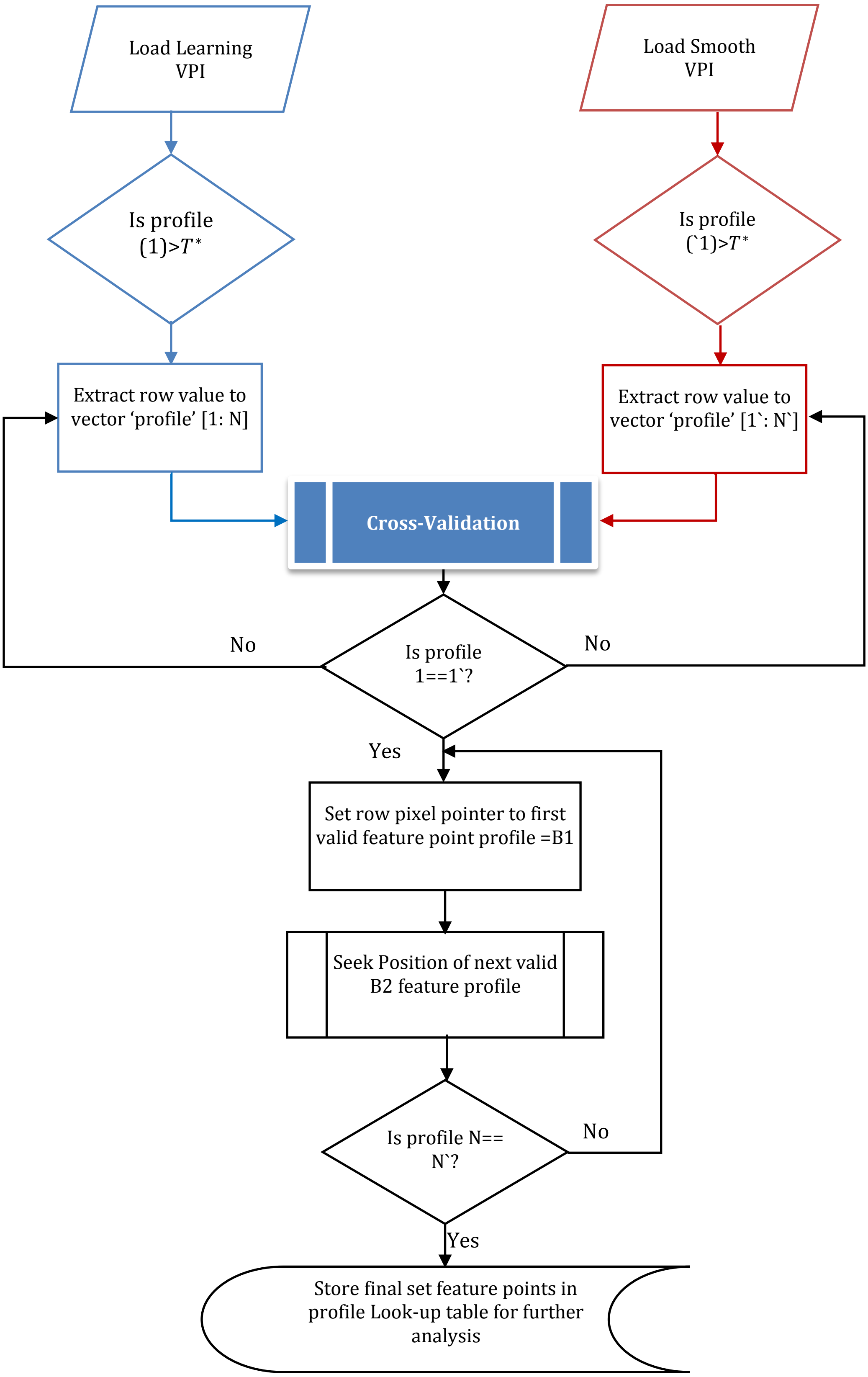

Load Learning VPI
Load Smooth VPI
Is profile (1)>T^*
Is profile (`1)>T^*
Extract row value to vector 'profile' [1: N]
Extract row value to vector 'profile' [1`: N`]
Cross-Validation
Is profile 1==1`?
No
No
Yes
Set row pixel pointer to first valid feature point profile =B1
Seek Position of next valid B2 feature profile
Is profile N== N`?
No
Yes
Store final set feature points in profile Look-up table for further analysis

Figure 4-3 Flowchart to illustrate the algorithmic design of generates FBP using optimal threshold.

When the highest local contrast is identified as the optimal threshold it is used to show the set of feature blocks on the central/reference 2DVPI. The feature blocks are obtained for additional analysis using the optimal FMS threshold process, where the spatial intensity distribution used as features to represent an image. The training task can be commenced once the training set is ready. If the intensity variance from a block is lower than a certain threshold value, then the block will be labelled as "untraceable". If the block contains enough features to enable a confident match (those where the value is greater than the threshold value of OT*) then it is labelled as a valid feature block.

The blocks classified as "untraceable" have low confidence and are discarded so that only blocks with an appropriate number of features are kept for further analysis. When performing model selection (feature selection), the data that has not yet been used can be split for use with fixed training feature points and validation sets. From this the generalization error confidence intervals can be determined from the cross-validation technique used with the feature point profile [144]. When the main goal is prediction, that is to determine how accurately a predictive model will perform in practice, this technique will be the primary one.

The following steps will explain how to evaluate a set of input features' performance:

1. Using the optimal threshold algorithm a subset X_s of the full set of input features $X = \{x_1, x_2, \dots \dots \dots, x_M\}$ is chosen for feature selection from the centre (learning) VPI.
2. A 3 × 3 median filter is used to reduce the noise from the learning VPI. As median filters are not based on needing values that are significantly different from those that are typical values in the neighbourhood, they can remove salt and pepper as well as impulse noise very well while details of the image are retained [145]. The way that the median filters work in successive image windows is comparable to that of linear filters [146].
3. From the features $\acute{X}$ =$\{\acute{x}_1, \acute{x}_2, \dots \dots, \acute{x}_M\}$ select a sub set $\acute{X}_s$ by using the optimal threshold algorithm with a smoothed VPI.
4. Unnecessary features need to be deleted accurately and to do this the generalization ability estimated by cross-validation is used as the selection criteria when optimizing performance by selecting features to estimate a feature points process and to

perform the evaluation of all the candidates in a feature subset. The candidate set of features $\acute{X}_s$ are set by the following:

$$\acute{\acute{X}}_s = \left\{ s \,\forall\, \acute{X}_s = X_s,\ s \in \{1,2,\ldots\ldots,M\} \right\} \tag{4.6}$$

Using this method it is possible to select only the valid feature point set that is present in both the initial leaning VPI and the smoothed VPI as the invalid features can be detected and removed without reducing the generalization ability. If the irrelevant features are maintained over-fitting may occur. It is also important to ignore features that have minimal or no effects on the output as the size of the approximator model should be kept small, which means the cross-validation of selected feature points is important as the aim is to determine the function between the input and output. There is a considerable reduction in the computation time, which is a benefit as when there are a large number of predictions the cost will increase significantly. Due to this the matching and corresponding process (the overall number of further steps) is reduced and a set feature point estimation can be optimized through modelling the piecewise nature of depth.

4.2.3 Depth Map Estimation Based on Auto Feature-Point

This section details the method for processing the entire H3DI and how to select the optimal feature block profile. The proposed depth map estimation from the H3DI system is summarized in Figure 4.4.

The methodology used to generate the 3D depth is demonstrated in the following steps:

1. Use the strong interval correlation between pixels that have been displaced by one micro-lens to extract the VPIs.
2. Then use the central VPI as the reference and training sample to compute the optimal threshold automatically from the distribution of variances in non-overlapping blocks.
3. Use fast bi-cubic interpolation to up-sample the extracted low resolution VPIs to produce the high-resolution information image [143].
4. Implement the Hybrid Multi-Baseline Algorithm (HMBA) that is a modified multi-baseline technique combined with an adaptive aggregation window, from Chapter 3, to estimate the depth map.

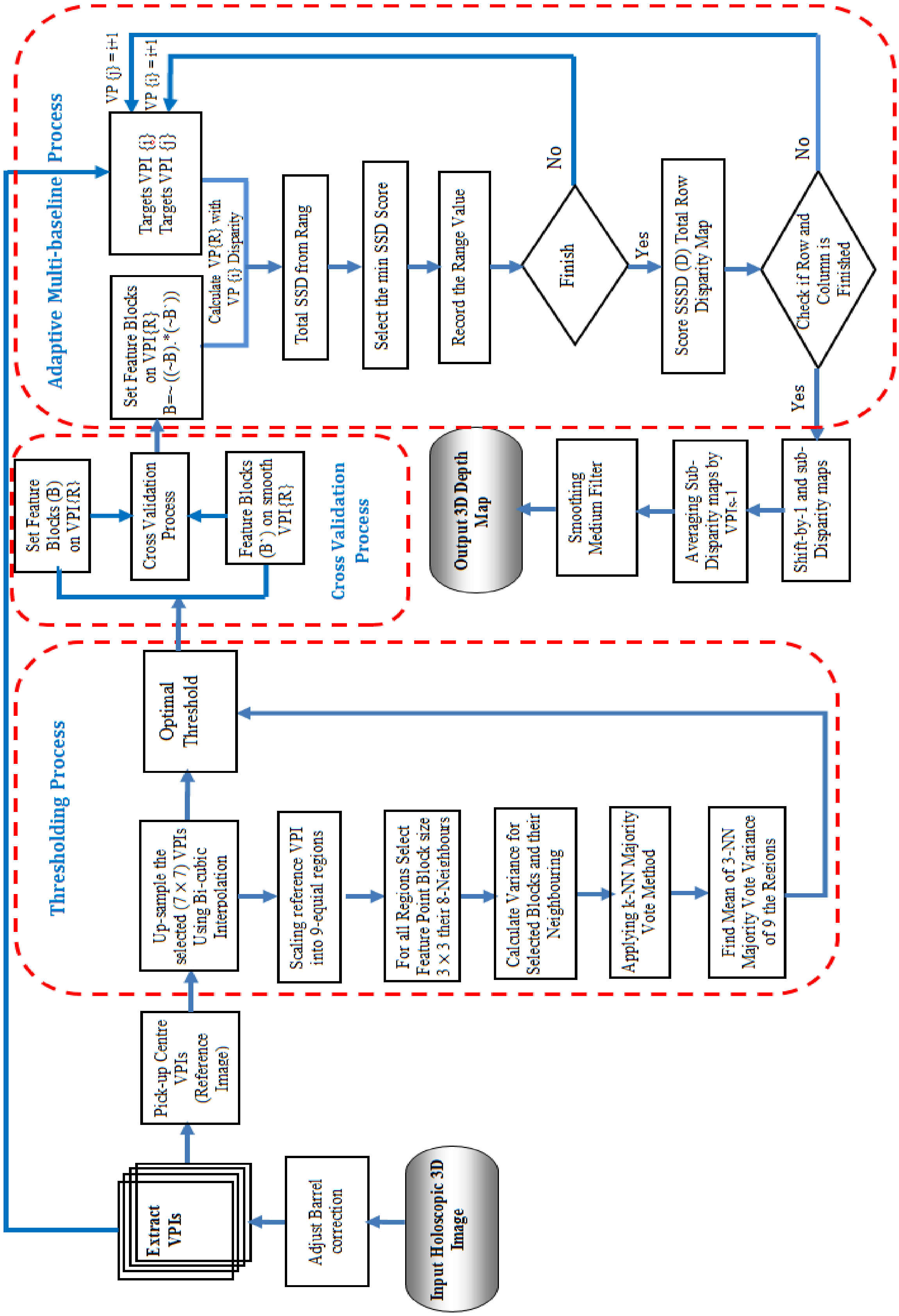

Adaptive Multi-baseline Process
VP {j} = i+1
VP {i} = i+1
Targets VPI {i}
Targets VPI {j}
Set Feature Blocks on VPI{R} B=~((~B).*(~B`))
Calculate VP{R} with VP {i} Disparity
Total SSD from Rang
Select the min SSD Score
Record the Range Value
Finish
No
Yes
Score SSSD (D) Total Row Disparity Map
Check if Row and Column is Finished
No
Yes
Set Feature Blocks (B) on VPI{R}
Cross Validation Process
Feature Blocks (B`) on smooth VPI{R}
Cross Validation Process
Output 3D Depth Map
Smoothing Medium Filter
Averaging Sub-Disparity maps by VPIs-1
Shift-by-1 and sub-Disparity maps
Thresholding Process
Optimal Threshold
Up-sample the selected (7 × 7) VPIs Using Bi-cubic Interpolation
Scaling reference VPI into 9-equal regions
For all Regions Select Feature Point Block size 3 × 3 their 8-Neighbours
Calculate Variance for Selected Blocks and their Neighbouring
Applying k-NN Majority Vote Method
Find Mean of 3-NN Majority Vote Variance of 9 the Regions
Pick-up Centre VPIs (Reference Image)
Extract VPIs
Adjust Barrel correction
Input Holoscopic 3D Image

Figure 4-4: Proposed flowchart depth map estimation from OH3DI using 7 × 7 low-resolution VPIs.

The non-informative descriptors, which either contain homogeneous areas or no feature points, can be discarded once the local set of feature descriptors for the whole VPI have started to be computed using the feature points setting and detection technique. Then, a matching and corresponding process is used between the remaining informative pixels of a reference VPI, which is selected to be the central one, with pixels of the target VPIs. For the information pixels from the reference VPI a set is formed that includes all the strong links with the target VPI's pixels. For each correspondence pair, they provide the coordinates of their own 3D origin, and when a set has pixels with the same correspondence pairs they are considered to be a chain. Each chain corresponds to an anchor point. The chance that false positive feature points may be formed is removed by making the correspondence set start from successive VPIs. Then Eq. (3.16) is used to compute the final 3D coordinates of the feature point.

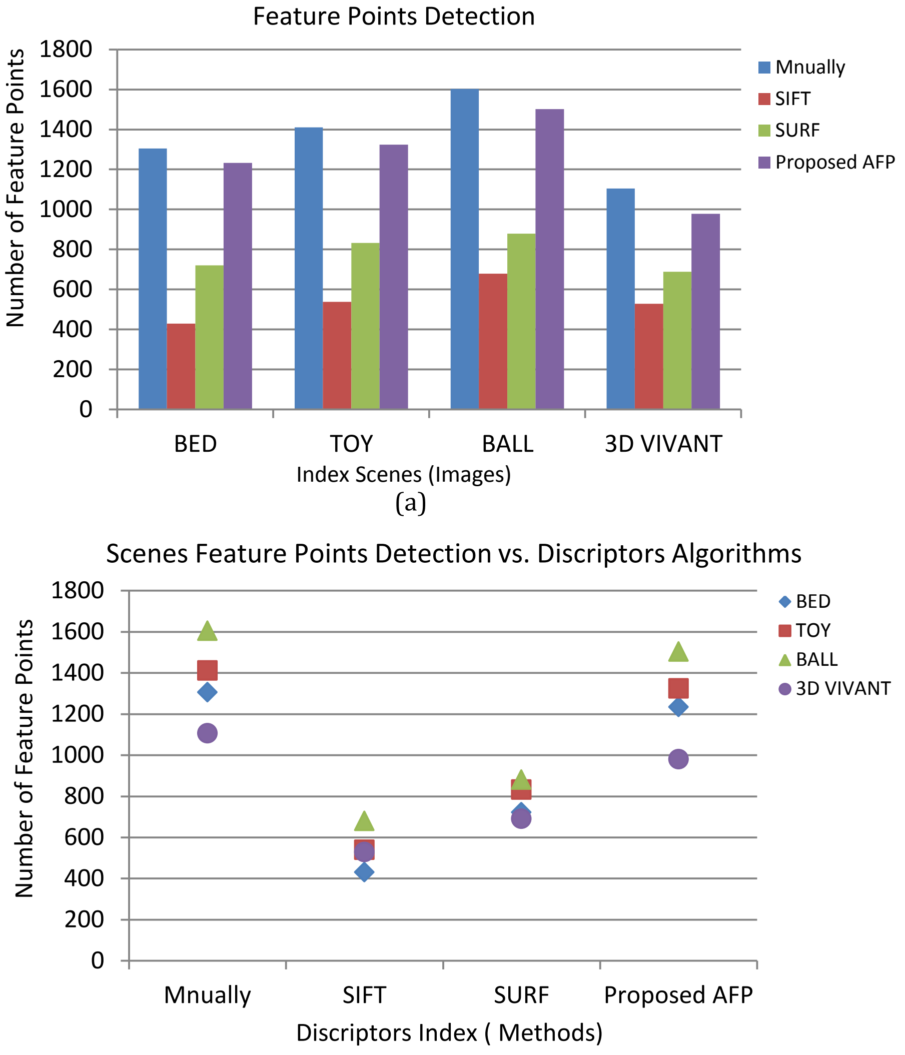
4.2.4 Evaluation of Auto Feature-Point (AFP) Detector

In this work a synthetic database was created in which the ground truth was known, this enabled the quantitative evaluation of the proposed algorithm and its components' performance by measuring its efficiency and intrinsic variations. The parameters of the database were included in Chapter 3. The commonly used tool for setting and extracting distinctive feature points for image matching is Scale-Invariant Feature Transform (SIFT) [147, 148] and Speeded Up Robust Features (SURF) [149-150] descriptors. These local descriptors were used as comparisons to evaluate the performance of the auto feature point's detection efficiency. The comparison methods are based on the computation of statistics related to the pixel intensity or colour values in the area that contains the points of interest. This means that the local image patches are discretized or quantized in discrete bins by the descriptors, which means that every time pixels are sampled for computing the bins they are unique each time. Both processes follow a similar approach but in SURF the computational cost has been reduced. In addition, a number of real-world UH3DIs and OH3DIs were used to test the accuracy and efficiency of the method when estimating the depth. The results from the feature points detection for the four detectors used (manual threshold , SIFT, SURF and the proposed AFP detector) are contained in Table 4.1 for the "Bed", “Toy” “Ball” and “3D VIVANT” data sets. The densest clouds of interest feature points were obtained by the manual and AFP detectors.

It can be seen from Figure 4.5(a, b) that for the proposed AFP detector its interest feature points were close to those of the manual threshold method. The Mean Relative Error (MRE) was used to evaluate the performance of the proposed AFP algorithm and has been defined in Eq. (3.20). The proposed AFP detector was able to achieve these impressive results due to the cross-validation technique that ignored input features that had only a small effect. Table 4.1 demonstrates the average and mean relative error of the detected feature points from four employed algorithms. It shows the ability of the proposed AFP descriptor algorithm to produce an enormous number of reliable feature points in comparison to SIFT and SURF. Experimentally, notable improvement results were achieved on overall depth estimation maps especially on low-textured and texture-less images, when the AFP algorithm produced a large number of feature point descriptors. The proposed AFP was competent, and created an average of 1260 feature points on each synthetic image, which was close to the number from the manual threshold algorithm (1356). Whereas, the widely-used feature point descriptors tools (SIFT and SURF) detected on average 544 and 780 points, respectively. The proposed AFP algorithm being able to produces a large number of reliable feature points, which helps on the overall depth estimation, especially on images poor in gradients and texture. However, the proposed AFP is more proficient in producing feature points on average per image from both types of H3DI system, whilst a minor increase of error on the proposed AFP algorithm has been noted in comparison with 4.26% and 4.07% on SIFT and SURF, respectively. However, the average feature points number of AFP were close to those of the manual threshold, since the highest percent error of 6.24% was observed for depth estimation maps within manual threshold as compared to 4.88% within the proposed AFP algorithm . This is due to the capability of the proposed algorithm to detect and identify a high reliable feature points. This means, that the proposed algorithm provides enormous quantity of feature points with high-quality representation. This means, that the most important information from the interest feature point can be captured. This has significantly led to an improvement in the performance of adaptive multi-baseline algorithm in terms of quality and speed. The underlying structure can be determined even when captured under different conditions, such as geometric or photometric transformation.

Table 4-1: Comparison feature points detected from different descriptors.

Synthetics Database	Manual	SIFT	SURF	Proposed AFP	Type of H3DIS
BED	1305	430	720	1233	UH3DIs
TOY	1411	539	832	1324	UH3DIs
BALL	1603	679	879	1502	UH3DIs
3D VIVANT	1105	528	689	979	OH3DIs
Average feature Point Descriptors	1356	544	780	1260	
Depth Mean Relative Error %	6.54	4.26	4.07	4.88	



(a)

(b)

Figure 4-5 (a) Number of interest feature points detected in four data set VPIs (Bed, Toy, Ball and 3D VIVANT, respectively in order). (b) Comparison of feature points detection from different scenes vs. descriptors algorithms.

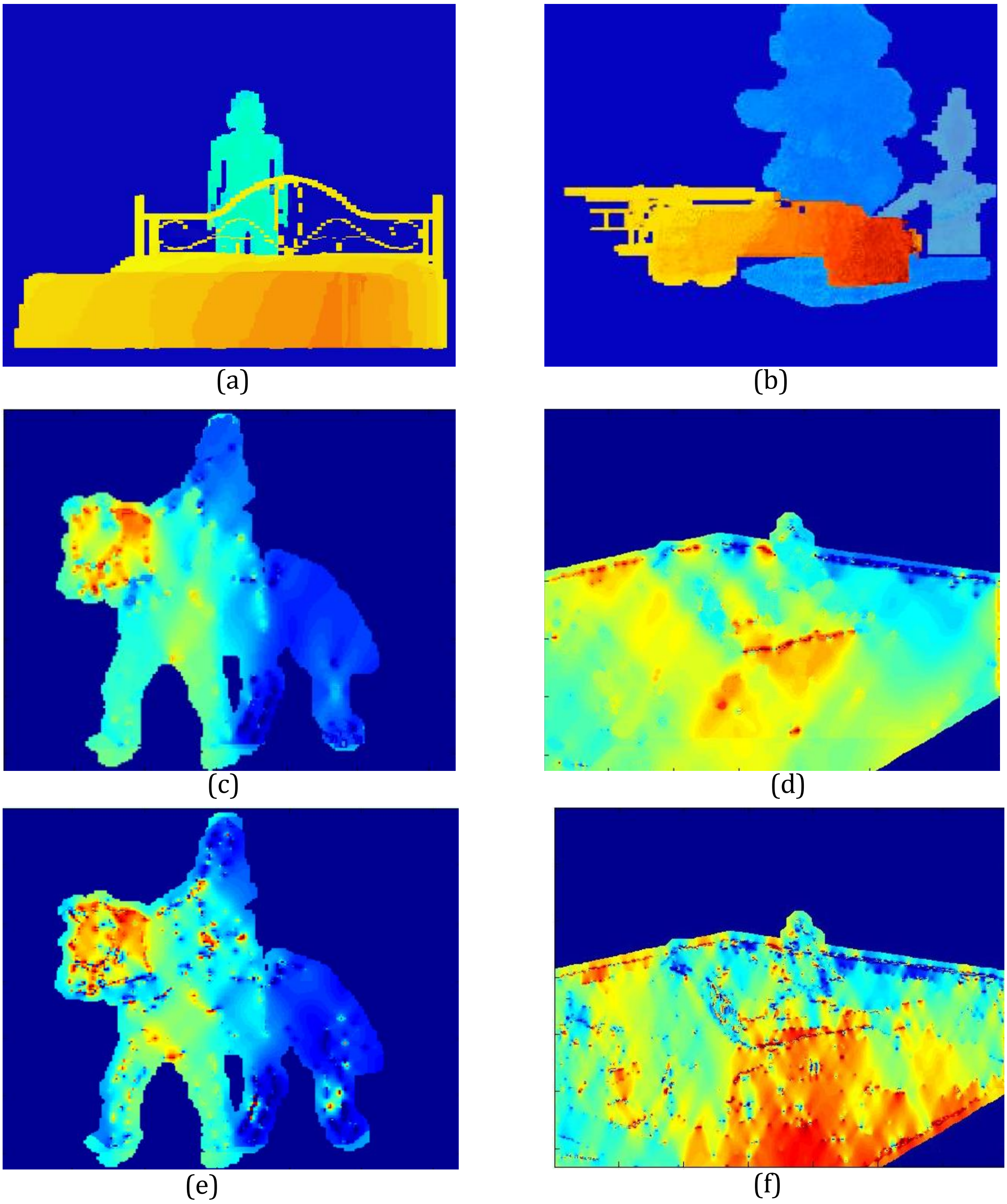
The results showed that, independent of the image, the manual method always obtains the maximum number of detections possible. A drawback of this is that the detection points are very close and can form clusters, which may cause problem when estimating the depth map. The AFP method obtained a similar number of detection points as the manual method but they are spread more evenly over the image. The SURF approach generated less detections than when using SIFT. A dense collection of detections are preferred as the detectors usefulness can be hampered if only a small number of detections are obtained. The number of feature points detected is not the be all and end all of a successful detector: their ability to overcome geometric or photometric transformations is also important.

4.2.5 Experimental Results

This section reports the experiments conducted to test the AFP algorithm. Two aspects related to it were the focus: i) efficiency: can it select and extract a number feature points that represents the number of vertices in the 3D depth map and ii) accuracy: precision loss after the simplification of the model i.e. comparison with the manually determined feature block selection and extraction threshold. For each experiment, the evaluation was carried out by comparing the model of depth estimation containing the complete set of features (full model) with one that had been manually simplified. A median filter was applied to smooth the estimated depth map. A median window size filter of 13 × 13 was calculated by systematizing neighbouring pixels so that they were in a numerical order and then substituting the middle pixel for the pixel that was being considered.

I. UH3DI Results

Figure 4.6 shows the results obtained from the proposed algorithm applied to a range of H3DIs that are both synthetic and real-world images. For all the images, the proposed method identifies the different objects correctly and differentiated them from neighbours, via the estimated depth values. However, contours and other fine details are not always included. This occurs most often for small objects or those with parts that are textured as there are only small differences between them and the surrounding parts. The VPIs are low resolution images, so for some of them, even the



(a) (b)

(c) (d)

(e) (f)

Figure 4-6 Proposed algorithm results show the effectiveness of cross-validation technique (a, b) synthetics 3D images, “Bed” and “Toy”, (c, d) real-word “Horseman” and “Tank” with cross-validation and (e, f) without cross-validation process. In all results the matching window size=17.

human eye finds it difficult to be able to determine the differences. However, despite these problems in all cases the method was able to make an acceptable approximation of the object contour as the specific type of object. In small baseline configurations the problem of occlusion is generally regarded as negligible, but as the baseline gets wider occlusion becomes a major problem that limits the depth map’s quality in the experimental results. Occlusion becomes a problem because it is possible for each pixel from the reference image to be hidden in many supporting images. The 3D structure of the scene induces the occlusion, and this cannot be determined until the correspondence is established, which is the final aim of the algorithm (Figure 4.6 (e, f)). How the depth estimation and the baseline are related is revealed by the depth equation. To enable a precise estimation of the distance a long baseline is required. However, this causes a problem as when there is a long baseline, a much larger disparity range must be considered when identifying the matching position. This means that it is more likely to generate a false match. It has been proven that ambiguity is reduced and matching precision increased when multiple-baseline matching cost functions are used in stereo with multiple camera images [126]. This work modified the algorithm from [126] to generate a new chain of target pixels from the reference VPI pixels as shown in Figure 4.7 that have the same matching feature points and corresponding process. From this the occlusions can be reduced as VPI pairs are used to reduce the width of the baseline.

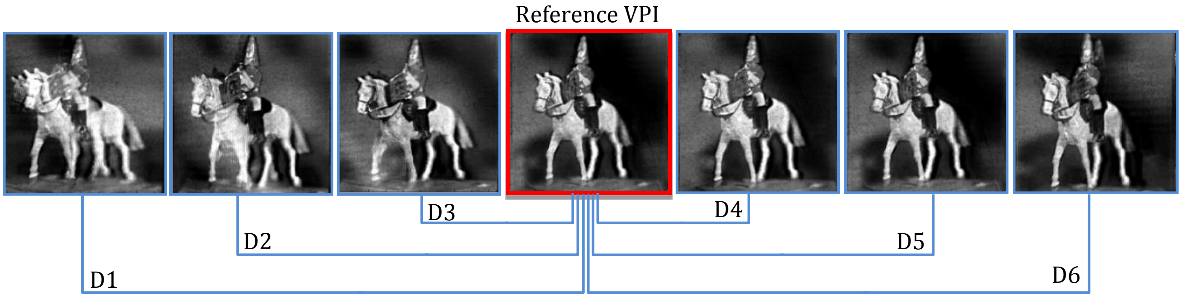
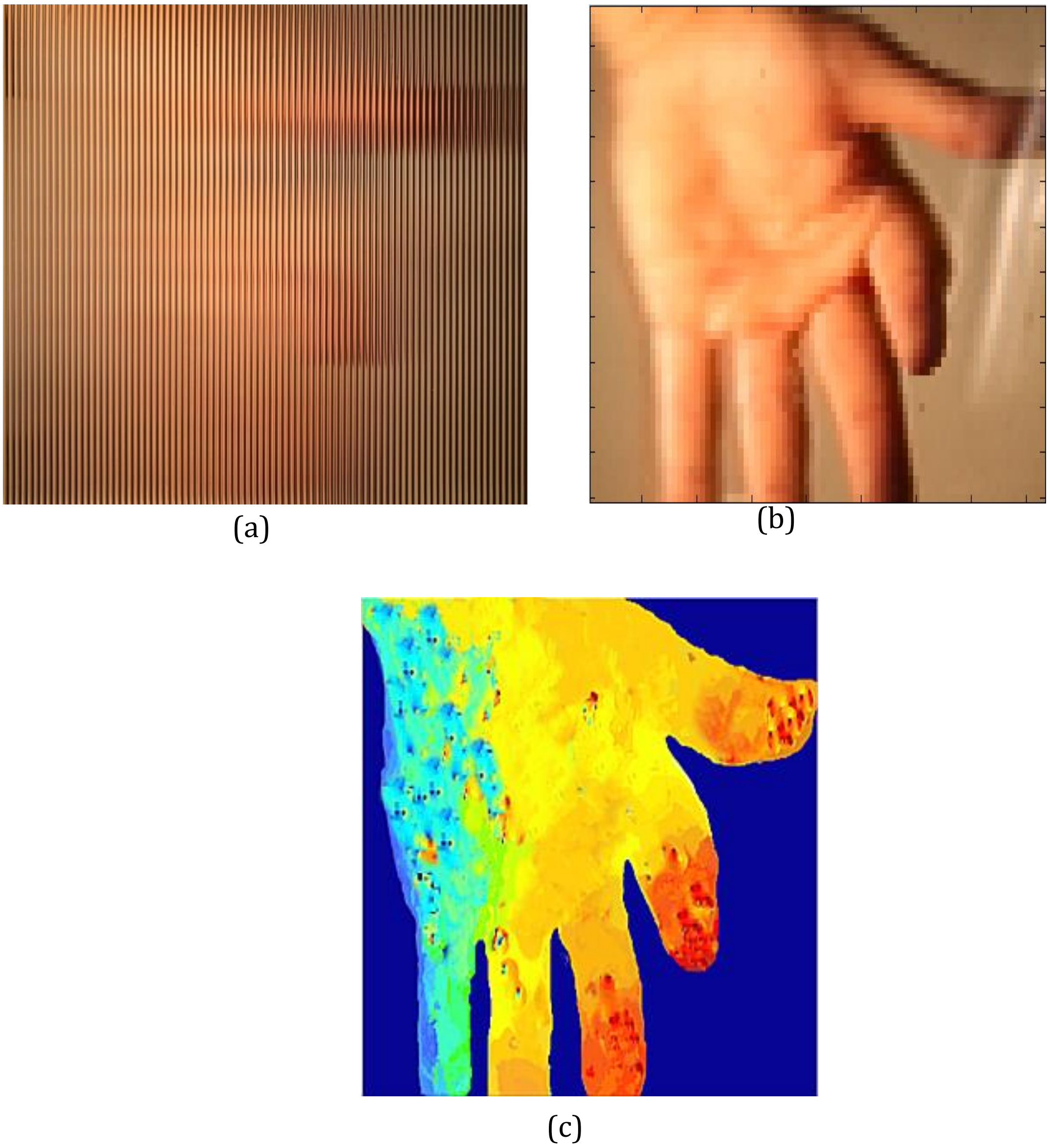


Figure 4-7 The chain corresponds to pixels from the reference viewpoint image among the target viewpoint images, where D1, D2, ... , D6, are the disparities between canter (reference) viewpoint image and targets.

The image shown in Figure 4.8(a) is another real-world H3DI that was captured by the 3D VIVANT PROJECT using a cylindrical lens array that was different from the one used to capture the “Horseman” and “Tank” images. In the micro-lens array each of the 84 cylindrical lenses was 67 pixels across, which generated 67 VPIs at a pitch size of 1.65 mm with a focal length of (f) = 2 mm. The “Palm” image is 5628 × 3744 pixels. For this the EIs were 67 × 3744 pixels and the VPIs were 84 × 3744 pixels. There is a lot of noise in the extracted VPIs and they have a low resolution: the range of depth values is very narrow (Figure 4.8(b)) and the palm seems to be very flat. It should be noted that the algorithm produced a smooth surface and was able to show the depth transitions between each finger (Figure 4.8(c)). In this situation the algorithm took a long time as there were 21 VPIs that required the up-sampling of 84 × 3744 pixels 21 times to obtain the depth map from the 21 × 21 matching windows.



(a)

(b)

(c)

Figure 4-8 Results on real-world "Palm". (a) Palm UH3DI, (b) up-sampled VPI, and (c) estimated depth map.

II. OH3DI Results

A simple experiment carried out at 3D VIVANT lab was performed on the real-world OH3DI named "Box-Tags" to evaluate the efficiency of the approach strategy that was used in the adaptive multiple baseline technique. The image is formed from a box placed in front of the camera and Tags separated by 200 mm that start 600 mm away from the camera and go up to 3000 mm. The image contains multiple depth planes as the distance between each Tag is invariant. The box was placed at a 45-degree slant from the Tags (Figure 4.9(a)) and its thickness was measured in the experiments with the camera starting at a distance of 600 mm from the ”Box”. In this experiment the diagonal of the rectangular parallelepiped is used to represent the real depth of the box. The image capture process used a 21-megapixel image size and a 5D Canon camera with a 50 mm main lens. To enable the distance between the main lens image plane and the micro-lenses as well as the micro-lenses to the image sensor the main lens is attached to the camera using a mountable extension tube to provide a flexible way of adjusting the distance. For the micro-lens the pitch size is 0.9 mm while the focal length is 0.25 mm. Due to the micro-lenses being square the main lens aperture was modified to also be square to enable the sensor space to be used more effectively. As shown in Figure 4.9(b) two local areas were chosen that showed the foreground “box” region at tag 1000 mm (*p1*) while the other was the background at tag 1200 mm (*p2*) in the reference VPI that enabled the depth to be measured to corresponding matching positions in the two widows in the multiple pairs of VPIs.

For the centre array (7 × 7) of the VPIs a series of correspondence pixels were detected in the 29 × 29 extracted VPIs array system at a resolution of 193 × 129 pixels for the “Box-Tags” image. The depth measurements when using the proposed technique on the “Box-Tags” image are shown in Table 4.2. The algorithm’s efficiency was evaluated using the MRE (Eq. (3.20) Chapter 3). The box’s real length can be determined as 200 mm through the use of a calliper gauge and geometric projections. The depth of the Box can be estimated using the compensated expected disparity as identified from the position when the global minimum is reached by the score functions. As illustrated, when the box is at the 1000 mm tag position the matching position (disparity) of 46.6 for the red patch in the foreground and 18.56 for the yellow patch in the background from searching range from -50 to +50. This translates into the box’s surfaces which give the depth at 338.225 mm and 134.71 mm, respectively, which gives a thickness of 203.5 mm. The error is less than 2% when the result is compared with the real thickness of 200 mm.

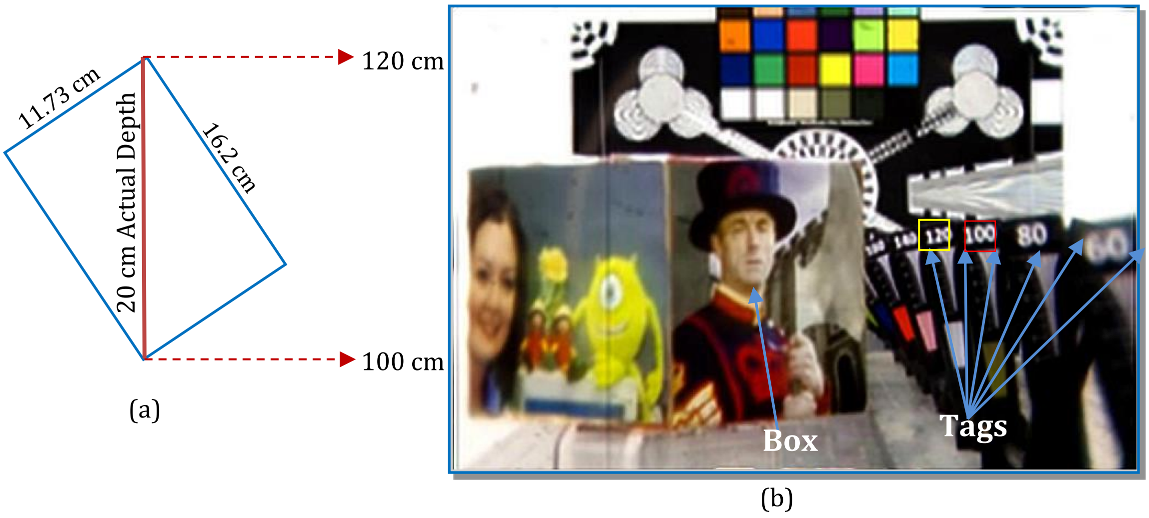


Figure 4-9 “Box-Tags” image captured in 3DVIVANT PROGECT lab. (a) The measure of the actual “Box” depth, and (b) display reference VPI of Box-Tags, the blue arrow points to the box and tags, the red and yellow are two windows chosen for depth estimation from the foreground and background region planes, respectively.

Table 4-2: Depth measurement results on real-world “Box-Tags” image.

Camera Distances /mm × 10	Estimated Depth/ mm × 10	Actual Depth/ mm × 10	MRE (%)
60	20.3	20	1.5
80	20.32	20	1.6
100	20.35	20	1.75
120	20.38	20	1.9
140	20.6	20	3
160	20.85	20	4.25
180	21.45	20	6.75
200	21.85	20	9.25
220	22.54	20	12.7
240	23.36	20	16.8

Figure 4.10 shows that for depths between 600-1200 mm the depth error was less than 2%. As the depth increased so did the error, but even at the furthest depth the error was only 17%. For the nearest depths with a 2% error, this is very good especially when compared to the first attempt with the CG-3DVIVANT OH3DI (Table 3.4, Chapter 3). Therefore, this technique could be used for the accurate and fast depth measurement.

The modified hybrid multiple baseline algorithm has had its effectiveness proven via the obvious improvement that can be found in the results when it is applied to both computer generated and real-world H3DIs, and this is especially true for regions that lack detail (e.g. tank). Figure 4.11 shows that for the "Box-Tags" and "Airplane-man" under laboratory conditions, the process detected the features and the depth that are correctly estimated. There is a marked improvement over previous results when considering the final depth maps that were generated form the available UH3DI and OH3DI data. The matching window size is bigger, the matching results are improved within the object/background region but the object has worse contours. Therefore, to improve the contour performance an edge detection technique was investigated for the setting and detection of feature blocks to identify the object contours.

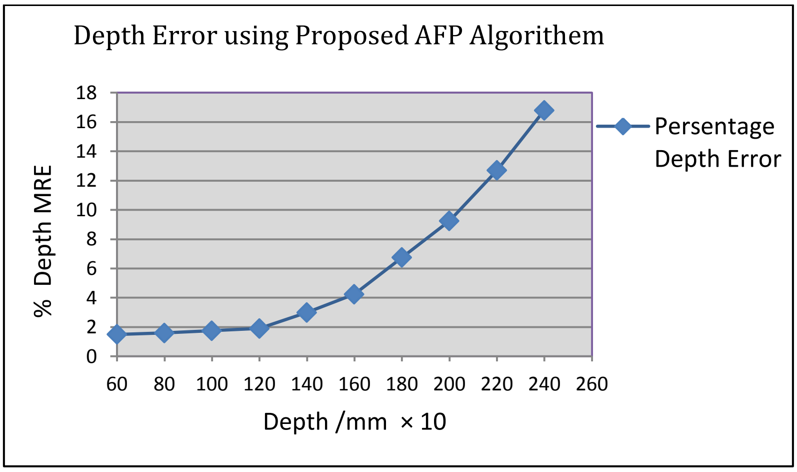


Figure 4-10 Plot experimental Box OH3DI error results from Table 4.2.

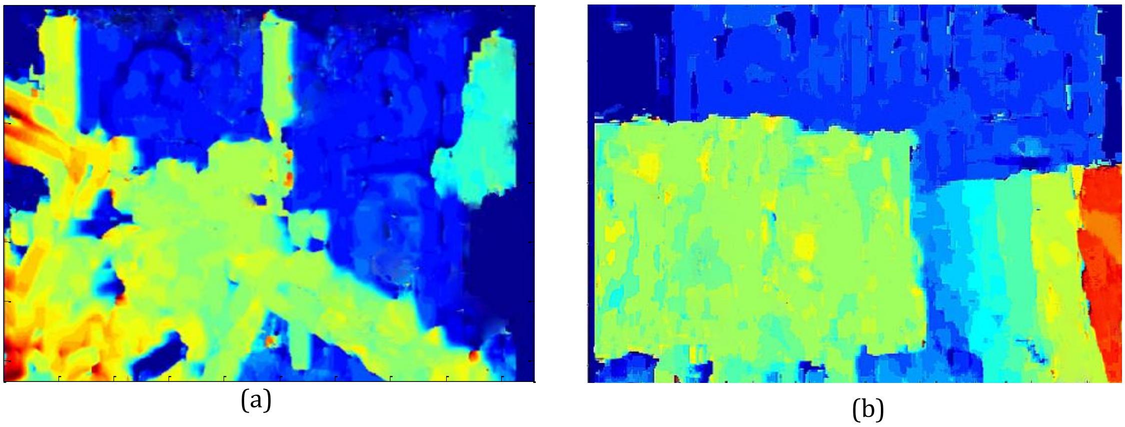


Figure 4-11 proposed approach final depth map results using matching window size 21 × 21 pixels. (a) Real-world OH3DI "Airplane-man" depth map and (b) "Box-Tags" depth map.

4.3 Feature-edge Detection Technique

4.3.1 Preface

Edges typically occur on the boundary between two different regions in the image, and therefore, are viewed as a contour across which the image intensity value undergoes a sharp variation that can be exploited to aid segmentation and 3D object recognition [151]. In other words, an edge is the boundary between an object and the background. The task of finding and localizing these changes in intensity is called edge detection. This is one of the most vital parts of early vision and is a prerequisite for automated vision systems. It is important for edge detection to be able to accurately find the edges and orientate them well. Edge detection maintains important information about the structure of the boundaries in an image to be processed while significantly reducing the amount of information to be processed. The three main stages of image smoothing, edge enhancement and edge extraction are included in most edge detectors [152]. In general, the edge detection has the image edges enhanced with a linear filter that approximates a first or second order derivative before a threshold is used in the edge extraction stage. The technique of thresholding is normally applied in the edge extraction stage as a post-processing step that separates real edges from noise. Most often a trial and error process is used to find the threshold value, which is then applied to the image to identify all the edges present. However, due to differences in the detector being used and the edge characteristics it is a taxing problem to find a threshold value that can be applied to various image types [153]. A major drawback in the thresholding based Feature-Edge Detection (FED) methods is the use of Automatic Feature-Edge (AFE) thresholding. The main problem relates to the automatic identification of a threshold value that is suitable for images with variations over a range of levels because of noise, multiple object with varying reflectance or illumination that is not uniform [154]. The threshold selection is very important for achieving good Feature-Edge (EF) performance, and automatic threshold edges have received attention.

To be able to produce a true volume that is accurate it is essential to model and implement an Auto-Feature-Edge (AFE) descriptor algorithm for a holoscopic system. The aim of using an AFE is to have a single detector that is able to integrate 3D cues and locate objects in a scene by detecting both contours (edge-based detection) and objects (region-based detection). For each FE block this will be an individual process of looking for large contrast changes. This work generates 3D-Interactive-Maps (3DIM) that include the use of depth estimation technique and FE segmentation into one technique automatically, and this is the novel aspect of this work. The algorithm was applied to two domains to establish its efficiency and robustness: a) for 3D depth map estimation setting and selection of the FE and b) separating the foreground objects from the background noise to segment the objects from the scene (see Chapter 6).

4.3.2 The Proposed Technique

To be able to create an accurate volume for a H3DI system it is necessary to develop a modelling AFE detector algorithm. This detects both the edge-based and region-based features so the image can be analysed efficiently by using the automatic threshold selection method. The FE points are the areas of the image where the values of the pixels are subject to large changes. There are three steps to be able to successfully set features via edge detection. Initially suppress any noise, before detecting and enhancing the edges, with the final stage being to localize and then accurately extract the feature points. The extracted feature points are then used to integrate the 3D cues that can be used to find the position of objects in the scene. Due to it being simple and fast to perform and easy to implement, thresholding is one of the simplest and most powerful methods to detect features [155]. A new technique for edge extraction has just been introduced that aims to provide the optimal trade-off between feature-edge localization and noise reduction. In this the objects and background from the centre (reference) VPI are separated into regions using an automatic bi-level (one threshold value) thresholding algorithm. However, when a single method is used with a wide range of image content, it is not always possible to generate strong feature edge information. To be able to identify distinctive FE blocks (large contrast changes) it is best to integrate more than one technique. The principle behind the use of the AFE algorithm in the Multi-Quantize Adaptive Local Histogram Analysis (MQALHA) algorithm is shown in Figure 4.12.

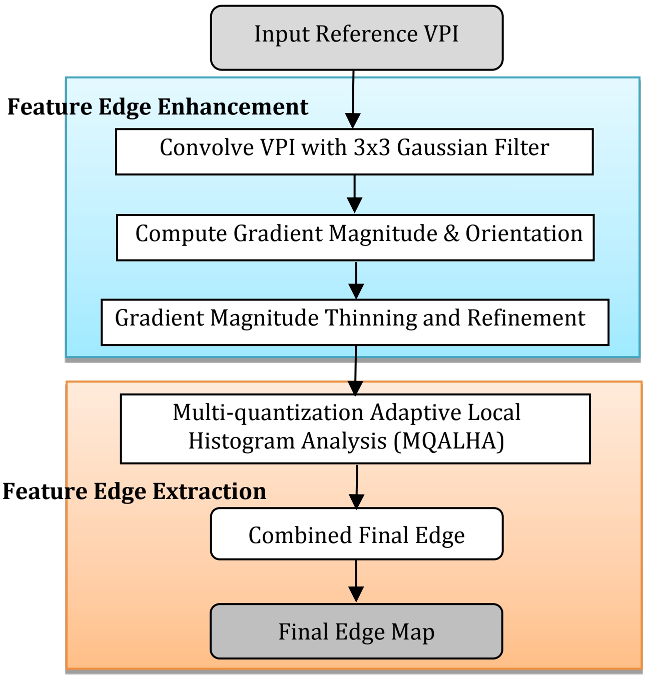

Input Reference VPI
Feature Edge Enhancement
Convolve VPI with 3x3 Gaussian Filter
Compute Gradient Magnitude & Orientation
Gradient Magnitude Thinning and Refinement
Multi-quantization Adaptive Local Histogram Analysis (MQALHA)
Feature Edge Extraction
Combined Final Edge
Final Edge Map

Figure 4-12 Enhancement and extraction of the Feature Edge phases in the AFE detector algorithm.

The processing can be divided into the following two major phases:

4.3.2.1 First Phase; Feature Edge Enhancement

First or second order derivative masks in the spatial domain are used in the edge enhancement phase to smooth the image and calculate the potential edge features. The following explains the main steps in this part of the overall AFE detector algorithm:

1. From the 2D Gaussian function: $G(x, y) = \frac{1}{2\pi\sigma^2}\exp[-(x^2 + y^2)/2\sigma^2]$, use a discrete Gaussian window = 3 to smooth the VPI to reduce the noise, where the width of the Gaussian defines the effective spread of the function is σ.
2. The main method to sort the good and bad edges is the gradient magnitude. The boundary between two regions increase as the contrast between the FE on a smoothed VPI ($\vec{x}$, σ) and the background is identified using the gradient magnitude. The magnitude $m\overrightarrow{(x)}$ and orientation $\theta\overrightarrow{(x)}$ of the greatest change in intensity are generated from each pixel's small neighbourhood [156] and derived from the difference between the pixels:

$$m\overrightarrow{(x)} = \sqrt{[VPI(x+1,y) - VPI(x-1,y)]^2 + [VPI(x,y-1) - VPI(x,y+1)]^2} \quad (4.7)$$

$$\theta(\vec{x}) = tan^{-1}[VPI(x,y-1) - VPI(x,y+1)/VPI(x+1,y) - VPI(x-1,y)] \quad (4.8)$$

Therefore, for the smoothed VPI the gradient can be written as:

$$VPI_{\sigma} = \nabla[G_{\sigma}(x) \times VPI(x)] == [\nabla G_{\sigma}](x) \times VPI(x) \quad (4.9)$$

3. Each contour should be thinned so it is only a single pixel wide. A non-maxima suppression process is performed to keep only FEs where the gradient magnitude is the local maximum. In the gradient magnitude VPI_{σ} all pixels are set to zero apart from those that have the maximum values, and these are not changed. To remove pixels with weak FE intensities each pixel's intensity is changed to be the average of the surrounding pixels in a 3 × 3 window.

4.3.2.2 Second Phase; Feature Edge Extraction

The reliable feature edges are selected in this phase by processing the computed gradient VPI. The aim in this step is to be able to increase the continuity of the FEs for the multi-resolution aspect to enhance the computational speed and to enhance the techniques efficiency so it can be used with a range of noise levels. The following lists the main steps:

1. Multi-quantization Adaptive Local Histogram Analysis (MQALHA) algorithm

The MQALHA algorithm is used in this step to increase the FEs' continuity, make the multi-resolution aspect faster and optimize the method for use with various noise levels. The quantization process, which is adaptive, is related to the image's estimated noise level. Figure 4.13 shows the MQALHA algorithm that was used to make and setup a new first derivative that was generated from the edge detection method so that it had the best integration of the detection and the edge's localization and resolution [157]. The task, which is multi-resolution, is performed using the MQALHA with local histogram analysis of the low and high edge maps that come from two quantized thinned gradient magnitude VPIs.

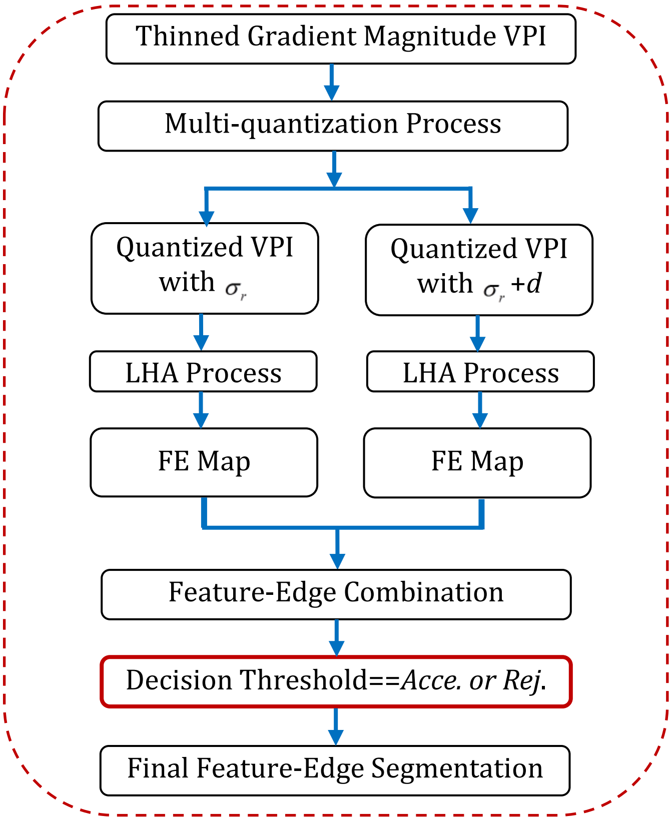

Thinned Gradient Magnitude VPI
Multi-quantization Process
Quantized VPI with σ_r
Quantized VPI with σ_r +*d*
LHA Process
LHA Process
FE Map
FE Map
Feature-Edge Combination
Decision Threshold==*Acce. or Rej.*
Final Feature-Edge Segmentation

Figure 4-13 MQALHA flow diagram for detecting and extracting feature-edges.

The following steps explain the algorithm:

First step: Image Quantization

Using the scheme presented in [157] and shown in Figure 4.14, the multi-quantization process scheme is used to automatically quantize the refined thinned gradient magnitude VPI into two 12-level intensities.

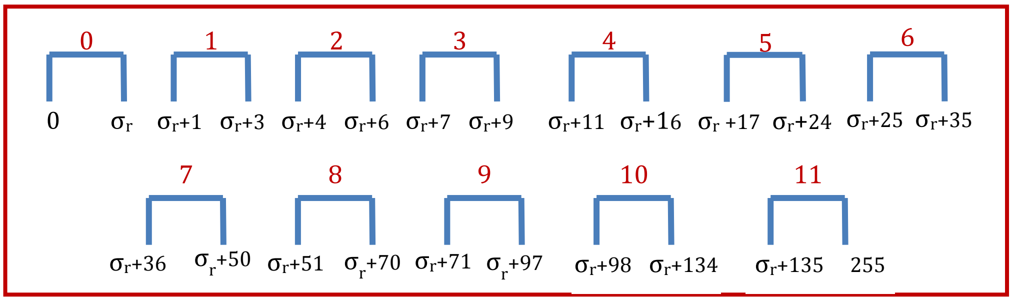


Figure 4-14 Adaptive 12-level non-uniform quantizer [157].

In the gradient magnitude histogram the small peak fluctuations were smoothed using a 1D Gaussian filter with a standard deviation of 0.8. The first local maximum in the smoothed histogram was located and the grey level was set to σ_r that corresponds to the position of the first maximum. The first quantized VPI is generated by using this estimated value to quantize the thinned gradient magnitude VPI. The σ_r value is moved a distance (d) right so that the second VPI is produced from a higher starting point than that in the 12-level quantizer. The value of estimated σ_r as the grey level that synchronization with a full-width-at-quarter-maximum (FWQM) height of the distribution is used to automatically predict the distance (d) and model the noise inhabitants in the gradient magnitude image.

In this work due to the VPIs having a lower resolution (from orthographic projection image) than the perspective projection images, the higher value of (σ_r+d) was experimentally selected to correspond to the position of a full-width-at-quarter-maximum (FWQM) of peaks in the noise distribution (see Figure 4.15 (a)). Due to the noise level in the image, the second quantized starting point will increase or decrease along with σ_r. The noise can be reduced by the adaptive quantization process at this stage, which is an advantage as this means that larger local histogram smoothing operators are not needed. This results in the overall edge extraction algorithm having a greater computational speed. To robustly represent the local histogram without losing information the quantization step is required, and it is needed to enable the processing to be conducted on a small block.

Second Step: Optimal Threshold

To extract the two edge maps at different resolutions local histogram smoothing and thresholding is undertaken on the quantized VPIs, which are divided into 4×4 non-overlapping blocks [155]. A 1-D Gaussian kernel of width W = 3 is used to smooth the blocks that are computed from the local histograms of the blocks. Then depending on the smoothed histogram's shape each block is classified as either a background block or EF block. Background blocks are those where the histogram is unimodal, and in these all the pixels are set to a value of zero. If the block is not in the background, take the smoothed histogram and differentiate it before taking the first valley's position as the threshold value for that block. For all pixels where the threshold value is exceeded by their quantized grey levels they are considered to be edge pixels and set to a value of one [157]. In the lower threshold FE map the thresholded block has the VPI of σ_r and is saved, while in the higher threshold FE map the thresholded block has the VPI of (σ_r+ d) and is saved for further processing. The auto threshold results are shown in Figure 4.15(b, c) for the "Horseman" VPI. More detailed explanation is available in Chapter 6 (section 6.4.1).

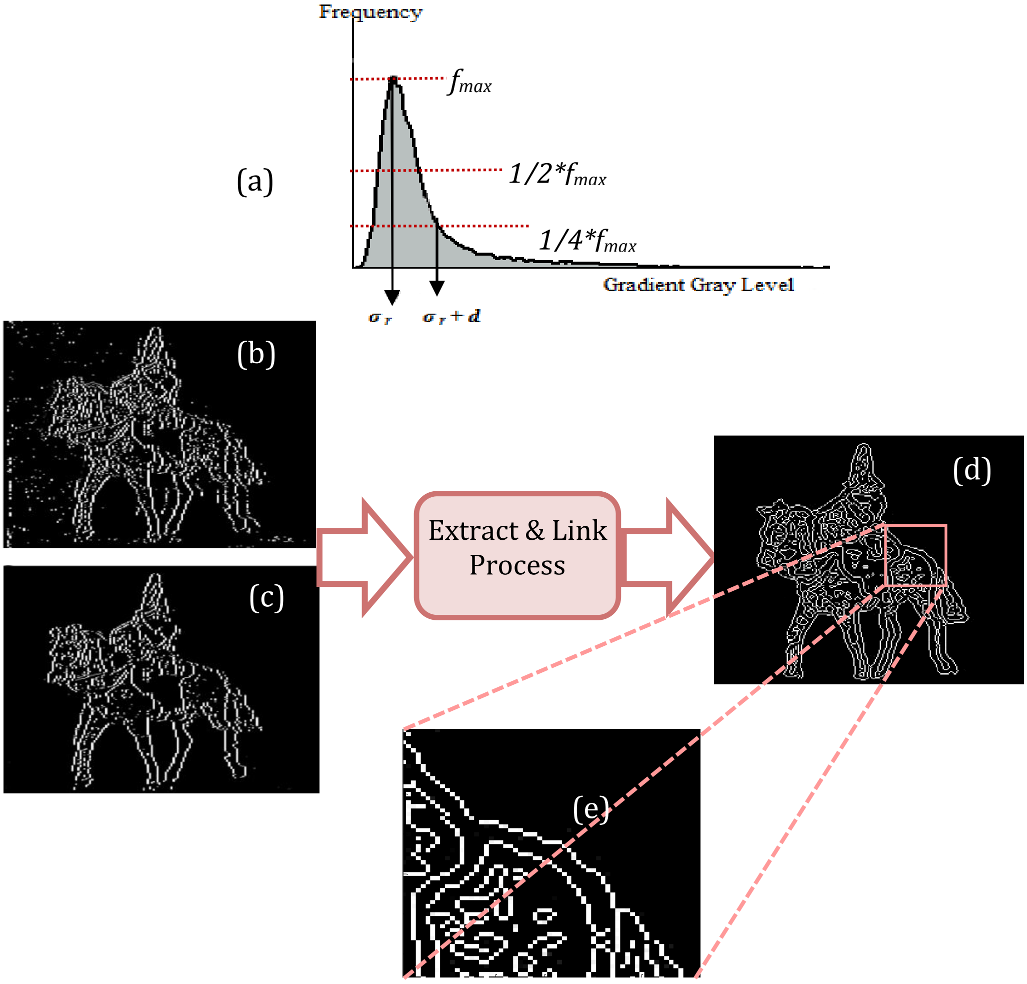


Figure 4-15 AFE extraction phase, a) first starting point ($\boldsymbol{\sigma_r}$) (low threshold) setting strong edges and second starting point ($\boldsymbol{\sigma_r}$+d) (high threshold) setting weak edges, b) low thresholding result on "Horseman" real-world VPI (noisy map), c) high thresholding resultant on "Horseman" (without noise), d) applying extraction and link process to set final EF map and e) magnified *section of* a final FE map.

2. Concatenating and Linking Edge Process (Combination Process)

This process combines all the edge maps obtained from the use of the MQALHA algorithm into a single edge image. The two EFs maps are combined via the edge combination process so that a "final edge map" is produced by extracting and linking the discontinuous segment FEs from the higher threshold and the lower threshold. An endpoint is found in the higher threshold FE map to start the process. Then using a local neighbourhoods' window size of 3 × 3 in the direction of the non-maxima suppression stage, the end points in the lower threshold FEs map are examined for possible connected successors. Figure 4.16 shows the process where the successor edge point is generated by considering the values of three probable neighbours from the lower threshold edge map that were determined from the endpoint's direction. Then if an edge is found in the lower threshold edge map, this will be extracted by the algorithm and linked to the endpoint in the high threshold map, and this new pixel will then become the endpoint.

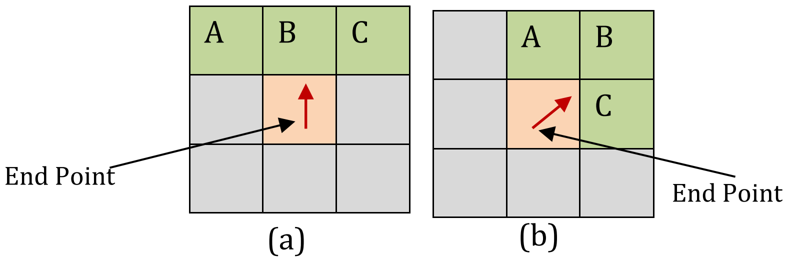


Figure 4-16 Probable successor location of the end point [157]. (a) The three possible successors pointed by 90° central pixel and (b) 45°.

The process will continue until all the connected edge points in the lower VPI are extracted and combined. This means that all the features from the higher VPI map can be extended and carried through to the final FE map (Figure 4.15(d, e)). Due to these features and depending on the task and object domain, the method for getting the required feature map and derivation of the final representation will be different between applications.

As the FE descriptors described above are scale invariant, illumination invariant and noise resistant, they are considered to be robust and efficient.

4.3.3 Depth Estimation Based on Feature-Edges

Using the method explained in section 4.3.2 for the estimation of a 3D object map from a multi-baseline this part will consider the contour based depth estimation problem. A flowchart of the adaptive hybrid depth map estimation technique that eliminates spurious edges in the UH3DI depth edge map via alternative implementation of feature edge detection is shown in Figure 4.17.

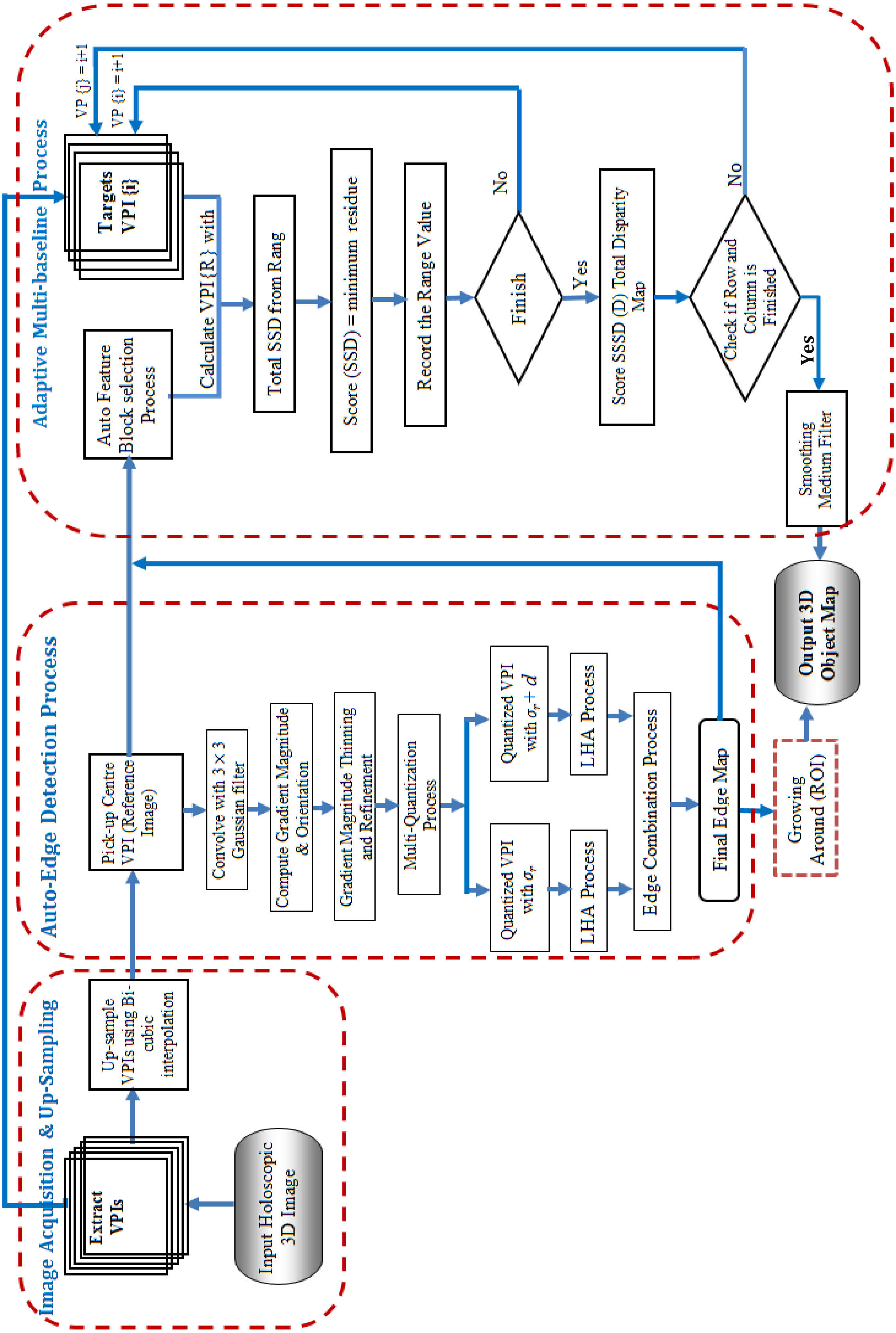


Image Acquisition & Up-Sampling
Input Holoscopic 3D Image
Extract VPIs
Up-sample VPIs using Bi-cubic interpolation
Auto-Edge Detection Process
Pick-up Centre VPI (Reference Image)
Convolve with 3 × 3 Gaussian filter
Compute Gradient Magnitude & Orientation
Gradient Magnitude Thinning and Refinement
Multi-Quantization Process
Quantized VPI with σ_r
Quantized VPI with $\sigma_r + d$
LHA Process
LHA Process
Edge Combination Process
Final Edge Map
Growing Around (ROI)
Output 3D Object Map
Adaptive Multi-baseline Process
Auto Feature Block selection Process
Targets VPI {i}
VP {i} = i+1
VP {j} = j+1
Calculate VPI {R} with
Total SSD from Rang
Score (SSD) = minimum residue
Record the Range Value
Finish
No
Yes
Score SSSD (D) Total Disparity Map
Check if Row and Column is Finished
No
Yes
Smoothing Medium Filter

Figure 4-17 Proposed depth map estimation and segmentation flow chart using MQALHA based on FED from UH3DI.

4.3.4 Experimental Results

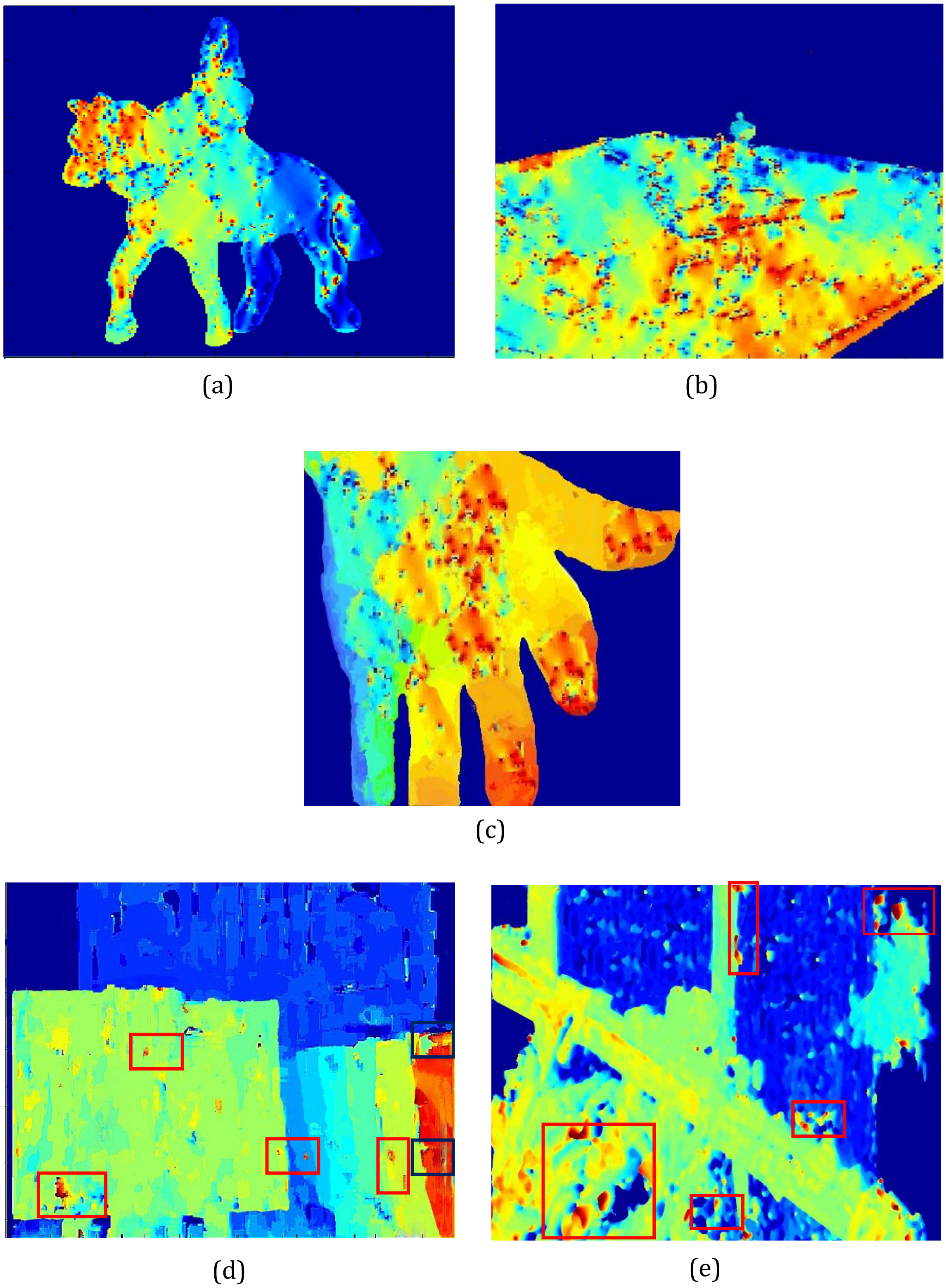
Figures 4.18(a-c) shows the results of the qualitative edge depth map on the “Horseman” real-world noisy scenario UH3DI, the “Tank” image that is lacking detail and the “Palm” that is a very flat image. Then the real-world OH3DIs “Box-Tags” and “Airplane-man” were used to test the generalisation algorithm with objects that had more complex shapes, as shown in Figures 4.18(d, e). The AFE algorithm was shown to be suitable for reconstructing the shape of the objects; however, there were prominent errors in the foreground regions that were caused by a lack of features for matching. The system can handle camera VPI changes, while also generating information about the 3D depth estimation and 2D foreground object segmentation. However, the thinning edge term (non-maximum suppression) caused some ambiguities related to the depth estimation map that prevented multiple responses. Due to the resolution being low in the VPI, not many details are visible and the feature detection and extraction are constrained. In addition, the score function judgment could be limited due to the illumination, focusing and similar patterns being present within the object scene.

Compared to the results produced using the AFP algorithm in section 4.2 the final depth maps from the AFE method have a marked improvement at the furthest depth ranges as well as having a visually satisfactory appearance. The AFE algorithm combines the adaptive multi-baseline technique producing results that show a clear improvement on the depth map contour segment. However, as far as accuracy is concerned the results from the AFP algorithm are much better than those from the AFE technique.

The new approach offered significant improvements in the reliability of the algorithm’s performance and improved the depth discontinuities at the boundaries of objects. However, due to ambiguities in the matching process wrong matches were shown in the qualitative depth image results, which were caused by the non-maximum suppression scheme that was part of the FED algorithm. However, the local geometric information, in which the associated image feature edge’s strength is indicated by a number, may be destroyed by a limitation of the non-maximum suppression. In addition, in weakly textured regions many valuable candidates may be over-suppressed by the non-maximum suppression [158] when a feature has its strength reset to zero, i.e., if it is not a local maximum its candidacy will be eliminated. For example, the condition: Strength $(b) >$ Strength (p) $\forall p \in O_b \{b\}$ where, O_b is a small neighbourhood of b (usually a 3×3 window cantered by (b)), is checked by non-maximum suppression to determine the candidacy of a feature block b [159]. The strength (b) is reset to zero if the condition is false. This can cause the number of possible neighbour blocks to be significantly reduced that are around the candidate feature block in the matching process. For the "Box-Tags" OH3DI results and depth estimation from two techniques the depth measurement efficiency was also involved. Table 4.3 includes the depth measurement results, while Figure 4.19 shows a comparison of the two techniques' performance, in which the estimated depth measurement error from implementing the AFE algorithm is less than the error percentage as determined using the AFP algorithm. However, despite the slight chance of false matches being generated, overall the algorithm gives very good depth position, which is due to the use of the adaptive multi-baseline algorithm. Both depth estimation techniques have had their effectiveness proven when generating accurate depth measurements via mathematical and experimental means. The promising outcome of the AFE depth estimation technique has been shown by the results where the 3D object depth segment contour has performed well with only minor errors that need to be improved in the future.

Table 4-3: Depth measurement results on real-world "Box-Tags" image using AFE algorithm.

Camera Distances /mm × 10	Estimated Depth/ mm × 10	Actual Depth/ mm × 10	MRE (%)
60	15.95	20	1.53
80	15.947	20	1.56
100	15.942	20	1.59
120	15.939	20	1.61
140	15.929	20	1.67
160	15.725	20	2.93
180	15.50	20	4.31
200	15.19	20	6.22
220	14.72	20	9.11
240	14.36	20	11.33



(a)

(b)

(c)

(d)

(e)

Figure 4-18 Proposed AFE depth map estimation and segmentation. (a-c) real-world UH3DI "Horseman, Tank and Palm", respectively, and (d, e) real-world OH3DI "Box-Tags and Airplane-man", respectively, where the red and black rectangles represents the errors.

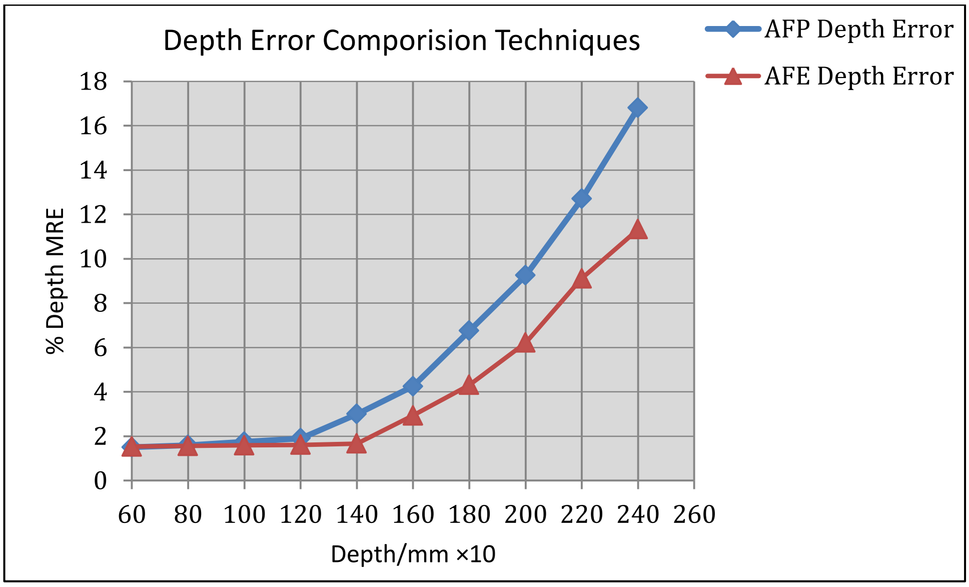


Figure 4-19 Plot of error results comparison from Table 4.2 and Table 4.3.

4.4 Summary

The information in this chapter was related to identifying algorithms that demonstrate the simplest way that H3DIs can have their depth estimated. The first and most important step towards determining the depth of extracted VPIs is to set and extract feature blocks. During feature selection the image is simplified so that it is in a form where it is easier to analyse and determine the positions of objects within it. This cannot be done manually as to do so would be prohibitively expensive and take too long for use in real-time applications. If a reconstruction process was to be used it would take too long to search the database that would be immense and this would affect the 3D model's quality. Therefore, it is better to automatically select distinguishing feature descriptors for the depth map estimation and then remove the least important ones. For use with a H3DI system, the automatic initialization techniques for creating and manipulating 3D models is fast and accurate.

The goal of this chapter was to set as well as detect 2D features that are present in the input VPIs and for which there are corresponding pixels in the set of VPIs. Then the 2D extracted features are used to compute the depth map by converting the features in them into 3D feature space. To enable the adequate production of descriptors, there must be features that can be easily recognised and matched reliably across the content of the set of VPIs. Two novel automatic local feature setting and extracting techniques were elaborated, which were built on binary data. The techniques were simple and widely applicable as well as being easy to compute. Often the features are the points or edges of the 3D object.

This chapter was presented in two sections with the first presenting a novel AFP algorithm that is capable of producing an accurate true volume of the object scene via extraction and setting of the unique features. The algorithm that is a local feature descriptor is based on the sub-dividing non-overlapping regions that are obtained from the VPIs and the distribution of sample variance in them. It was proven that the well-known SIFT and SURF descriptors were outperformed by the proposed AFP algorithm when setting and extracting the distinguished feature blocks. Using both real-world UH3DI and OH3DI data satisfactory quality depth maps were produced while the consistency of the measurements were proven in the OH3DI "Box",

The second section aimed to further refine the precision depth estimation map and presented an AFE descriptor algorithm that was based on verifying the importance of the objects' edges through adjustment of contours of the object based on edge detection to extract the distinctive feature edge blocks. This generates robust feature edge information that holds data about the shape boundaries of objects in the foreground that is closely related to the geometry of the 3D scene and background clutter is completely removed due to reflectance discontinuities. The efficacy of the approach depends on the robust estimation of edge features and their connection.

It is worth pointing out that, due to the object distances from the camera, the captured light rays for the object that is far enough away will likely hit nearly all micro-lenses. This results in representations of that object in nearly all EI's; therefore, the object will register with high spatial sampling. In other words, the same region of the object will be sampled in multiple EIs. As a result, the new (second) proposed algorithm was shown to be an improvement as at the nearest depth of 600-1200 mm it gave an error of less than 1.7% against the previous (first proposed algorithm) error of 1.9%. The AFE algorithm also showed improvement on the depth map contour segment that was due to using the adaptive multi-baseline technique. However, when the matching of the corresponding feature edge blocks was undertaken near untextured areas or those with similar texture patterns there were numerous ambiguities observed, such as the ambiguities observed on the flat image “Palm”.

Furthermore, for real-world H3DI the distinctive features obtained from the AFE algorithm are used to generate the foreground mask for 3D object segmentation that enables the location of objects that are in the foreground of the object plane. Therefore, for the foreground objects only the depth is calculated to separate the foreground object/objects in the scene from the background (see Chapter 6 for more details). From this it was possible to correctly estimate the relative positions of the objects. Regions that are textureless are likely to have the most prominent errors due to the thinning edge term (non-maximum suppression) that causes viable candidates to be suppressed when present in weakly textured regions.

It became obvious that a straight forward extracted VPI to represent the whole scene is a very limitative solution and thus leads to a limitation when extracting robust features. Enhancement of the VPI’s resolution is important to increase the depth estimation performance from the extracted low resolution VPIs to enable high-resolution image features to be produced that can be used to obtain super depth information maps.

CHAPTER 5

Depth Estimation Based on Super Resolution of H3DI

5.1 Introduction

The focus of this chapter is to use the orthographic VPIs approach as an alternative way to solve the resolution problem and improve the depth map's visual quality. This chapter aims to compensate for the poor features of the low resolution VPIs (LRVPIs) to develop a high-quality depth map of a 3D scene by generating super resolution multi-view images from a H3DI system. As previously stated, a H3DI system is attractive as it can store both intensity and directional information in a 2D image at fixed-time, which is modeling VPI. It is formed from parallel rays that are projected at different angles from the object. Accordingly, in practice images from a H3DI system have different properties than standard perspective images due technological reasons i.e. the radiance sensor resolution. In practice, the resolution of VPIs (orthographically projected rays) is much lower than the resolution of a traditional 2D image (perceptively projection rays). This is due to the similar technology of the radiance sensor that is used to capture both traditional 2D images and the multiple points of view present in the H3DIs. This results in a factorization of the VPI resolution, which is proportional to the number of VPIs in a H3DI that causes adverse image distortion. Therefore, further processing is required on the produced 3D-display image. To cope with this, in this chapter a new noval technique to reformatting a set of VPIs to a perspective view image is capable of generating a new form of VPIs in perspective projection with high resolution and wide Field Of View (FOV).

It is worth to pointing out that, the EIs of the H3DI system are vital for the estimation of High-resolution (HR) 3D information of the scene. The limit of the EIs resolution is due to the physical distance of the pixels (the pixel pitch or picture element) in the display panel. In other words, the absolute resolution of each picked-up EI should be severely decreased according to the dimension of an employed micro-lens array. Hence, it might be difficult to obtain the accurate depth data of 3-D objects from these low-resolution EIs. Therefore, the pixel pitch of display is a fundamental resource that rules the display quality of 3D depth or disparity maps. Therefore this technique's aim is to determine the full parallax super-resolution 3D depth map from a H3DI to generate 3D images that are more natural and fatigue-free.

5.1.1 Image Resolution

The amount of detail that can be observed from an image is the resolution. In cases where information needs to be extracted from images, the more detail there are the better it is to extract information from an image. Optical super resolution methods are expensive and are usually developed to enhance the resolution of an already expensive imaging system [160]. Henceforth, the term 'super-resolution (SR)' in this context is used exclusively to refer to the process of overcoming the sensor density limitation using signal processing methods. This means that super-resolution can be used to increase the sensor density of an image cheaply. By combining a number of images of the same scene additional information is added to the reproduction. In general, this information is high frequency content of the scene. Not all the LRVPIs are identical as there are some variations between them, for example, different viewing angles [161, 162]. A much better image of the object can be generated from the information that is in multiple LRVPIs as well as an understanding of the transformations between the VPIs.

5.1.2 Multi-Image Super Resolution Technique

Multi-image super resolution is a well-studied problem now and several methods by researchers introduced the idea of image super resolution techniques to reconstruct the SR image by receiving additional information from multiple LR images [163]. Then SR became an exciting research area and many SR techniques that are superior have been developed [164]. Most of the methods for multi-frame SR images from the literature can be divided into three stages as follows [164, 165]:

1. The first stage identifies the relative shifts between the LR images compared to the reference LR image using fractional pixel accuracy. Thus, the same coordinate system is used to register all images i.e. "*Registration Stage*".
2. To generate a uniformly spaced composite of the LR images to form the SR image, interpolation is used. There are several methods to perform the interpolation that has a key role in the framework. The nearest neighbour algorithm is the simplest interpolation algorithm, in which each pixel that is unknown is given the same intensity value as that of its neighbours. The drawback is that blocky appearing images are produced with this method.
3. In the final stage the blur and noise are removed from the SR image produced in the above stage "*Interpolation Stage*" i.e. "*Restoration Stage*".

Figure 5.1 represents the process of constructing a SR image from multiple registered LR images.

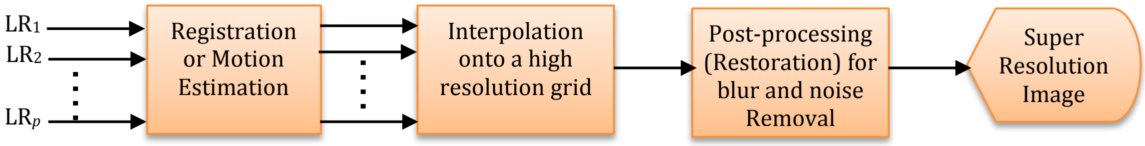

LR1
LR2
LRp
Registration or Motion Estimation
Interpolation onto a high resolution grid
Post-processing (Restoration) for blur and noise Removal
Super Resolution Image

Figure 5-1 Scheme for general super-resolution image generation from multiple low resolution images [164].

In the stages presented above, the assumption is that it is possible to map all pixels from all frames back to the reference frame using the motion contained in the vector information, from which the up-sampled frame can be obtained. The steps can be implemented either at the same time or individually depending on the reconstruction method chosen. This chapter's context is related to the SR of a light field, which is due to an extra amount of information about the scene being provided in each LRVPI. Therefore the aim is to produce an SR image that is clearer and more detailed from the 'fusing' of all the information rather than just duplicating the pixels to make a bigger sized image.

5.2 From Orthographic to Perspective Projection

The camera sensor's and micro-lens array's resolutions limit the H3DI system's performance. According to the sampling theorem, these define the trade-off between the recovered light field's angular and spatial resolutions [166]. In addition, because of diffraction, the micro-lenses' size restricts the sampling VPIs resolution [167]. As stated in Chapter 3, the VPIs are a collection of all the pixels at the same position in each EI and they have orthographic projection geometry. This means that a parallel grid without a vanishing point is used to sample the object space. The EI's field of view (FOV) is limited to $2\tan^{-1}(\varphi/2f)$, where the focal length and lens pitch are f and φ, respectively. The VPIs' resolution is not greater than the number of EIs as these are too small and coarse. This means that due to each EIs' limited size the images are low resolution. The EIs have the object's 3D information embedded as the set of EIs give the 3D object's ray space. In other words, the accuracy of the disparity detection has a direct effect on the quality of the generated VPI. Not many features (details) can be observed in the corresponding process as the resolutions of the VPIs are so low. Accordingly, it is vital to improve the quality of these images for more reliable feature correspondences through rendering SR or full resolution VPIs via transforming a set of VPIs into a perspective view.

At a particular plane a new interpolation technique is needed to generate the SR images from the H3DI system, therefore the new full resolution VPI rendering technique is proposed from unregistered sets of sample VPIs. Three stages are involved in the rendering of one plane in the high spatial resolution image. The hybrid interpolation algorithm overcomes the problem of artefacts, which are mainly produced in areas that are out of focus in the light field super-resolution method. The first is up-sampling a set of extracted VPIs, then shift and integrate the up-sampling set of VPIs and finally de-blurring or (restoration) stage. The method is presented in detail in Figure 5.2 that shows the stages of up-sampling, shift and integration of the orthographic projection LRVPIs that are for focusing a specific depth plane for any shift value to generate a SRVPI with perspective projection geometry.

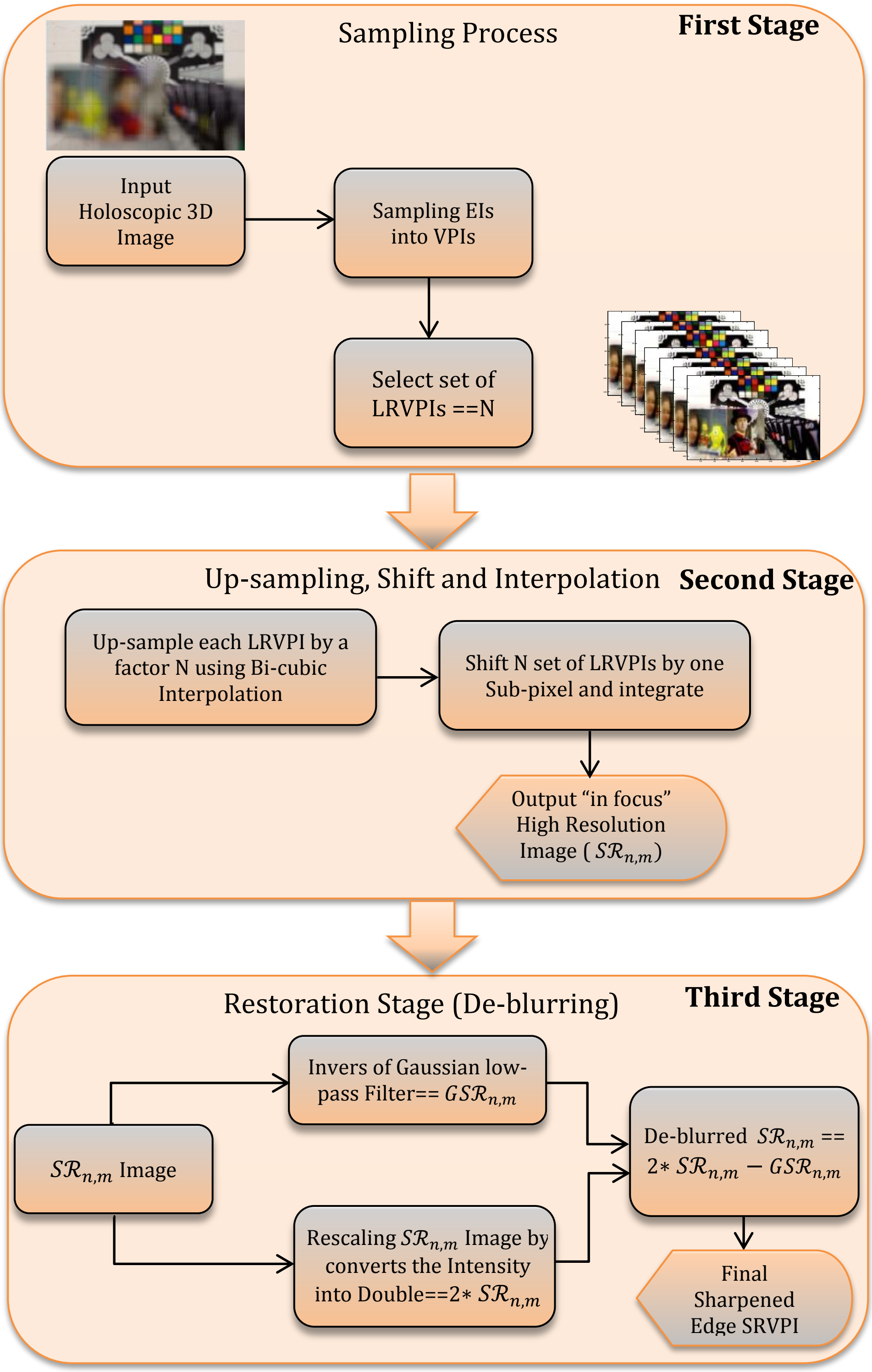

Sampling Process
First Stage
Input Holoscopic 3D Image
Sampling EIs into VPIs
Select set of LRVPIs ==N
Up-sampling, Shift and Interpolation
Second Stage
Up-sample each LRVPI by a factor N using Bi-cubic Interpolation
Shift N set of LRVPIs by one Sub-pixel and integrate
Output "in focus" High Resolution Image ($\mathcal{SR}_{n,m}$)
Restoration Stage (De-blurring)
Third Stage
Invers of Gaussian low-pass Filter== $G\mathcal{SR}_{n,m}$
$\mathcal{SR}_{n,m}$ Image
De-blurred $\mathcal{SR}_{n,m}$ == 2∗ $\mathcal{SR}_{n,m}$ − $G\mathcal{SR}_{n,m}$
Rescaling $\mathcal{SR}_{n,m}$ Image by converts the Intensity into Double==2∗ $\mathcal{SR}_{n,m}$
Final Sharpened Edge SRVPI

Figure 5-2 Flowchart of the proposed super-resolution VPI generated from H3DI system; displaying effectively the stages and processes of up-sampling, shift and integration of the N set of orthographic projection LRVPIs.

5.2.1 Up-Sampling using Integer Factor

During integer up-sampling zero-valued samples are placed between original samples using an integer factor to increase the image's sampling rate. When an integer factor is used to multiply the sample rate the undesired spectral images from the original image are added to that, and these have multiples of the original sampling rate at their center. This means that the up-sampled image is distorted compared to the original, and that interpolation digital filtering is vital to remove the undesired images. Often a simple but fast low pass interpolation technique is used for scene reconstruction and warping via high throughput machine vision tasks. Bilinear and bi-cubic interpolations are two such linear filters using pixel values for fitting locally linear and cubic functions, respectively [168]. In commercial software these two filters are the most common as well as simplest nearest neighbour interpolation.

Up-sampling is often needed for image matching and registration when using computer vision applications: it is often used to aesthetically manipulate images. In the context of this approach, the up-sampling occurs before the stages of shifting and integration. This allows the integration of sub-pixels from the same spatial point in different views and improves the image's resolution due to the representation of more pixels around the same point. However, due to the shifts between LRVPIs being arbitrary it is not always possible for the SRVPI to match an evenly spaced SR grid. This means that when there are non-uniformly spaced composite LRVPIs non-uniform interpolation is needed to generate a uniformly spaced SR image.

The following steps explain the up-sampling method used in this approach:

1. First one SRVPI is generated from a selected set of N number sampling LRVPIs. Identify each LRVPI as $VPI_{i,j,n,m}$, where the VPI coordinates are i, j and the parallel light rays of the OH3DI's coordinates are n and m. Figure 3.14 in Chapter 3 give the principles used to transform the EIs from the OH3DI system into VPIs.
2. Before applying the shift and integrations, each LRVPI should be up-sampled in the horizontal and vertical directions by a factor of N $(Ni \times Nj)$. A 4D stack of $VPI_{i,j,n,m}$ images is formed by placing adjacently the up-sampled VPIs horizontally and vertically from top to bottom direction. This step enables the enhancement of the

VPI's resolution, as the same spatial point from different views can be used for sub-pixel integration. Figure 5.3 shows the process of up-sampling using five EIs with three pixels for each.

The up-sampling process used MATLAB bi-cubic interpolation filter that is smoother and has less interpolation artefacts.

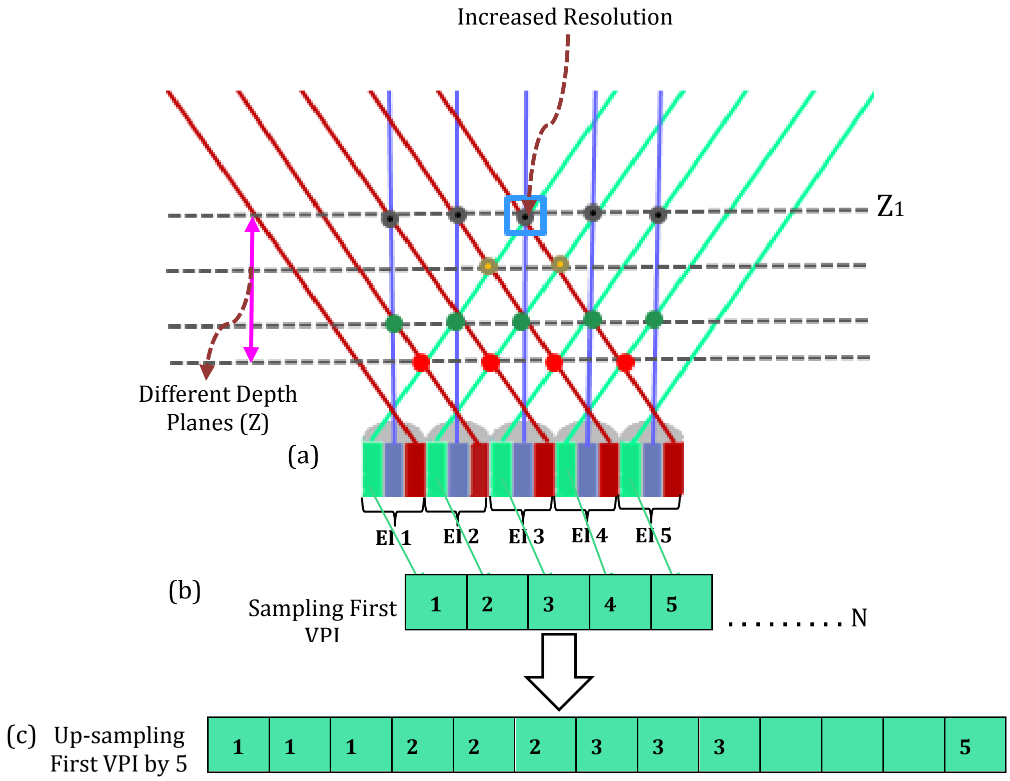


Figure 5-3 Example of up-sampling process, (a) 5 micro-lenses each with 3 pixels, i.e. 5 EIs each with 3 pixels, (b) sampling EIs into VPIs and (c) up-sampling by the number of the sets of VPIs (=3) using MATLAB quadratic interpolation.

5.2.2 Shift with Sub-pixel Precision and Integration

The human observer may be distracted or mislead by the fused image's quality if artefacts or inconsistencies are introduced. This might occur if the shift and integration (fusion) technique is not applied to make sure that the fused SR image has all the spatial and spectral information transferred from the input LRVPIs. Each LRVPI contains complementary information that means the maximum amount of information can be contained within the SR perspective projection geometry. Sub-pixel shifts are used to introduce the non-redundant information that is present in the LRVPIs. These sub-pixel shifts may occur due to uncontrolled motions between the imaging system and scene, e.g., movements of objects, or due to controlled motions [170]. The first work related to SR was about the shift and aliasing properties in the Fourier transform [171]. In frequency domain approaches the image observation model that can be handled is very restricted as real problems are more complicated. The spatial domain is mainly used by current researchers to address the problem as it is flexible and can model all image degradations. A non-uniform interpolation using a generalized multi-channel sampling theorem has been proposed for a set of spatially shifted LR images in which the LRIs are merged over a fine grid and a higher spatial sampling rate blurred image is obtained [170]. A good overview of existing SR methods is given by *Macwan et al*. [172].

This work has considered moving many LR images at the sub-pixel level (fractions of a pixel) and then positioning them in a spatial multiplexed manner as a way to enhance the perceived resolution. This is closely linked to the SR techniques in [173] and [174] where a group of LR images that have different translational offsets are used to compute a high-resolution image. Once the LRVPIs have been shifted with precision at the sub-pixel level, the new information contained in each VPI can be exploited to obtain a SR image. When shifting is done at integer units each VPI will contain the same information as previously and there will be no new information to reconstruct the SR images. This means that SR reconstruction is possible if sub-pixel shifts can be done between LRVPIs.

An "in-focus" high spatial resolution VPI plane can be produced when the window size is returned to where the plane is focused after shift and integration where one sub-pixel = ($1/N$) from the input set of the LRVPIs (N). Only one plane of high spatial resolution image will be generated from this process. As well as the all-in-focus image other depth planes are needed which means different shift values are required.

From the integrated up-sampled N set of LRVPIs ($\mathcal{K}$) the SRVPI ($\mathcal{SR}_{n,m}$) can be generated as follows:

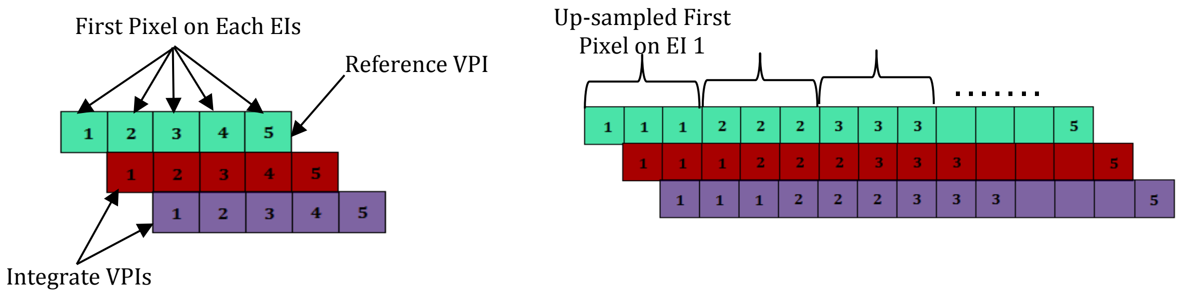
$$\mathcal{SR}_{n,m} = \sum_{k=1}^{N} \sum_{l=1}^{N} \mathcal{K}_{n+\delta_x(1+i),i,\ \ m+\delta_y(1+j),j} \tag{5.1}$$

where, the up-sampled VPI coordinates set are n and m; the VPIs indexed numbers from 1 to N are k and l. For the horizontal and vertical directions the sub-pixel shift parameters are δ_x, δ_y, respectively, for the Nth LRVPI with respect to a reference image. These represent the modified index of the n and m coordinates. The numbers of EIs' in horizontal and vertical resolutions are i and j. As long as the shifts are known in this proposed technique's context, then $\delta_x = \delta_y = 1$. The shift parameter's value δ is given by:

$$\delta = \frac{pixel\, size}{N} \tag{5.2}$$

The process is shown in Figure 5.3(a) with an example where the depth plane Z1 can be observed by using a shift value of 1 from different EIs. This means that for pixels under EI n*shift from all the different EIs will be collected at point *Z1*, in which n=1, 2,…, *N* represents the number of EIs. From this, a single depth plane (*Z1*) will be "in-focus" with one sub-pixel shift using up-sampling-shifting and integration algorithm.

As shown in the blue square, the FOV is directly increased via the enhancement at the depth plane *Z1* and the rays from the neighbouring VPIs. At the depth plane *Z1* three pixels were intersected that increased the resolution in the refocusing process. It is worthwhile pointing out that, SR reconstruction is not possible from images shifted by an integer amount as they would not contain any new information (only the same intensity values at the same spatial locations). For example, when a reference VPI that has been shifted by one pixel is integrated with another VPI the modified pixel = two sub-pixels, which, as shown in Figure 5.4(a), results in the same resolution as if the VPIs had not been shifted, meaning there is no resolution enhancement. Using Eq. 5.2, when undertaking a shift of *1/N* (e.g. half a pixel, if two VPIs are involved) the pixel shift = one sub-pixel, in this situation new information will be contained in each sub-pixel LRVPI, which can be used to generate a SRVPI. Figure 5.4(b) shows the relative shifts between LRVPIs compared to the reference LRVPI that can be estimated with fractional pixel accuracy.



(a) Shift and integrate with out up-sampling

(b) Sub-pixel Shift and integrate process

Figure 5-4 Relative shifts between LRVPIs compared to the reference LRVPI, (a) without up-sampling using 5 micro-lenses i.e. 5 EIs each with 3 pixels, (b) resolution enhancement, where, up-sampled all VPIs before the implementation of shift and integration process.

5.2.3 Post-Processing

Due to over or under fitting during the reconstruction of SR images associated with the fixed shifting of the neighbouring LRVPIs there is often blurring. Images that are blurred do not have definable features [175]. Therefore, a new simplified model of the basics for a typical de-blurring filtering process is employed as a point-spread function. This filter is the inverse of the Gaussian blur filter, in which there are two processes: smoothing and sharpening. The third stage of Figure 5.2 shows the process that is explained in the following:

1. Smoothing Phase: The blurred $S\mathcal{R}_{n,m}$ VPI is convoluted with a 2D Gaussian blurring low-pass filter kernel in each direction (to the nearest odd integer) so that noise and small functions are suppressed. The mask is calculated with a Gaussian function. Within the filtering process both the standard deviation (σ) and size of the mask are important. The image will be blurred slightly when there is a small sigma and mask size, while a large sigma and mask will result in heavy blurring. The most general 1D directional Gaussian function used when the mask is calculated is the following:

$$G\ (x) = \frac{1}{\sigma\ \sqrt{2\Pi}} e^{-\frac{(x^2 - a^2)}{2\sigma^2}} \tag{5.3}$$

where, (σ) is the standard deviation and (a) is the statistical expectation that causes the distribution shifting along the *x* axis so it is zero: a = 0. This means that a zero mean Gaussian distribution is assumed for the distribution. The following gives the horizontal and vertical 2D Gaussian distributions:

$$G\ (x, y) = \frac{1}{\sigma\ \sqrt{2\Pi}} e^{-\frac{x^2 + y^2}{2\sigma^2}} \tag{5.4}$$

Then, the mask M of size $b \times b$ is given as:

$$M\ (x, y) = G\ \left(x - \frac{(b+1)}{2\sigma}\right) \times\ G\ \left(y - \frac{(b+1)}{2\sigma}\right) \tag{5.5}$$

where*: 1<=x<=b and 1<=y<=b.*

The 2D Gaussian filter is applied on the original image as a point-spread function. This is then split into two 1D distributions that are in the horizontal and vertical directions. This means, firstly determine the 1D mask weights and then use them as filters for every line as a 1D signal then filter every filtered image column as a 1D signal. To smooth the blurred SRVPI its error is modelled as a Gaussian noise filter that is used to smooth it. For each signal element $\mathcal{SR}_i$, the Gaussian filter's convolved image signal is $\mathcal{SR}_{n,m}$ = $\{\mathcal{SR}_i\}$, and for each EI the new modified $\mathcal{SR}'_i$ will be calculated. This means that each EI in the center of mask (M) of size $b \times b$ will be in the centre of the window. The new filtered value is given by the following, which multiplies each element in the mask by its corresponding weight and provides a sum of all the products:

$$\mathcal{SR}'_{n,m} = \ G\ (x, y) *\ \mathcal{SR}_{n,m} \tag{5.6}$$

This causes the low frequencies in the element to be amplified while suppressing the high frequencies.

2. Sharpening Phase: The higher frequencies in the image are amplified and the noise is not via the use of image sharpening methods [176]. While successfully creating sharper images, the methods also have the possibility of causing haloing artefacts. This approach is able to improve the image without generating artefacts, since the low frequencies are suppressed while the high ones are amplified. This is done by rescaling the non-filtered $\mathcal{SR}_{n,m}$ by converting the intensity to double precision, which is subtracted from the filter $\mathcal{SR}'_{n,m}$. This means that, it sharpens edge

preservation of Gaussian blurred image with the same parameters, and it works as follows:

$$\mathcal{SR}''_{n,m} = 2* \mathcal{SR}_{n,m} - \mathcal{SR}'_{n,m} \tag{5.7}$$

where, the modified original signal element is (2* $\mathcal{SR}_{n,m}$) which when subtracted from the smooth signal element is $\mathcal{SR}'_{n,m}$, and hence the de-blurred $\mathcal{SR}''_{n,m}$ image is produced. This has an inverse smoothing effect in the horizontal and vertical directions and is a simple and effective way to ensure that the signal has had the noise removed while ensuring the detail remains in the SR image. This gives gentle smoothing and maintains the sharp edges by suppress the low frequencies and amplifying the high frequencies depending on the smoothness of the signal.

This de-blurring filter technique is a simple yet very effective process that suppresses noise while maintaining the SR image's details by avoiding the over-sharpening effects while also removing the blur noise. This means that this interpolation based adaptive filtering approach is suitable for real-time implementation. This has a prominent impact on the images that have been refocused as the structural details in the focused regions of the image were enhanced.

5.3 Implementation and Experiments

Using UH3DI and OH3DI (two types of real-world image), this section presents the results of experiments showing the new method's reliability and efficiency. Data invariant linear filters will be used for comparisons as they are popular in commercial software. The well-known Bi-cubic (MATLAB precision) filter is among them, furthermore, the frequency domain algorithms by *Vandewalle et al.* [177] are used for comparison. This is a graphical user interface (GUI) for the MALAB program used to implement multiple image registrations and for the reconstruction algorithms found in SR images. It will also be compared with *Keren et al.'s* [178] spatial domain method that was based on Taylor expansions.

5.3.1 UH3DI Experimental Results

This section presents a range of SRVPIs that were created with the algorithm described above and then compared with other techniques. The results obtained from the "Palm" UH3DI are shown in Figure 5.5, which had a pitch size of 1.65 mm, focal length of 2 mm and each micro-lens had 67 pixels at a VPI resolution of 84 × 3744 pixels. The proposed SR technique uses a set of 5 LRVPIs {$V_{1,} V_{2,} V_{3,} V_{4,} V_5$}, i.e. ($N$ = 5) to produce one SRVPI. This separately used a bi-cubic interpolation algorithm that has the properties of low computational complexity, smoother and fewer artefacts. This means that the VPI size becomes 420 × 3744 pixels when a sample rate integer factor of five is used. Each LRVPI is up-sampled in the horizontal direction five times. The sub-pixel number = 1 (the original image's pixel size which has the shift value of $1/N$) is used to shift all the up-sampled LRVPIs before they are fused. For the post-processing stage the parameters are [15 15] and 2 that give the Gaussian filter size and σ, respectively. This means that the FOV is increased directly for the fusion of rays for the neighbouring VPIs. Only the horizontal direction has the up-sampled shift, which integration algorithm used. This means that there is an eight pixel increase in the horizontal direction making the size of the de-blurred SRVPI equal to 428 × 3744 pixels. The following shows the calculation for the up-sampled- shift and integration of the SRVPI:

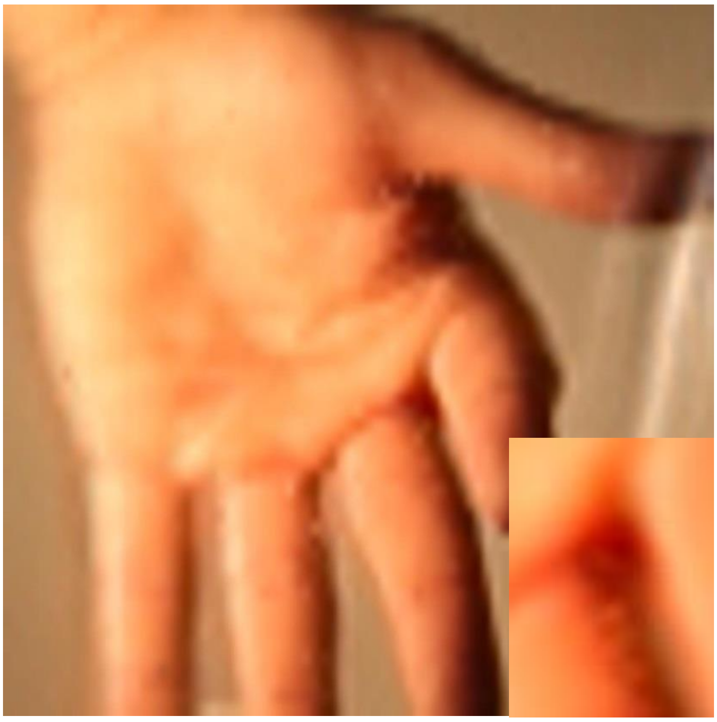
$$(VPI_i)_R * N + (N\text{-}1)\ (\delta_x + 1) = 84 \times 5 + (5-1)(1+1) = 428$$

When compared with the other algorithms used this new method showed significant improvement in the experimental results showing its effectiveness when using the same data throughout. Figure 5.5(a) shows the original reference LRVPI that has a resolution of 84 × 3744 pixels, the results from the new method are shown in Figure 5.5(b) and the comparison to other algorithms shown in Figures 5.5(c-d).

This work only used the judgments by human observers to test the subjective image quality as this is how the images will be judged in reality to prove the performance of the proposed algorithm when there is no-reference image/ground truth available. More details of the subjective quality assessment are available in Chapter 6 section 6.5.2. While the physical measurement of the images' quality i.e. objective image quality is used to evaluate the images so that the reconstructed SR images are compared with known original images (reference) that were used to obtained sequences of LR image.

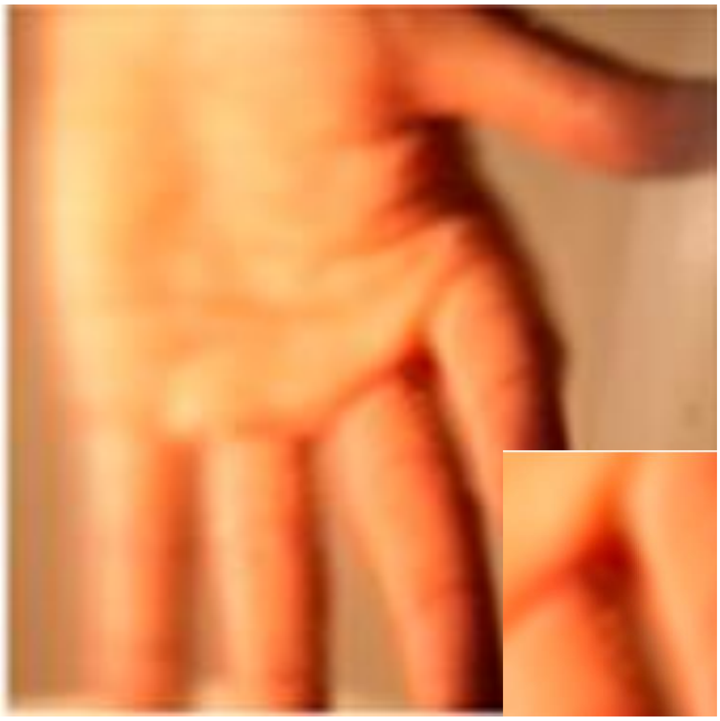
(a)

(b)



(c)

(d)



(e)

Figure 5-5 Simulation results of the different algorithms on “Palm” UH3DI images (flat, noisy image). (a) One of the five LRVPI size 84 × 3744 pixels used as reference for up-sampled shift and integration algorithm, (b) reconstructed SRVPI size 428 × 3744 pixels from the proposed algorithm, (c-e) reconstructed SRVPI size 420 × 3744 pixels from [177], [178] and bi-cubic MATLAB software, respectively.

There are many parameters that determine the optimal number of LR images that are needed to perform up-sampling-shifting and integration algorithm, such as the accuracy of the sampled set of LRVPIs, imaging model and total frequency content, etc. Intuitively, there is a trade-off between the number of images and level of reconstruction: better reconstruction can be obtained with more images while at some point a limit will be reached on the level of improvement obtained. Sometimes even with many images it is not possible to get a sharp SR image. The resolving power is limited by the blur, noise and inaccuracies from the signal model. The result of using 18 LRVPIs (N = 18) with the new method is shown in Figure 5.6 and it can be seen that the image is over-smoothed. This shows that it is only really practical to increase the resolution by a factor of five sets of VPIs. In this work, the VPIs are sampled from H3DI system; therefore, there is no existence of the original VPI image (see Chapter 3). The aliasing artefacts present in the input VPI, Figure 5.5(a), are removed in (b) using the sampled VPIs approach to form the SRVPI.

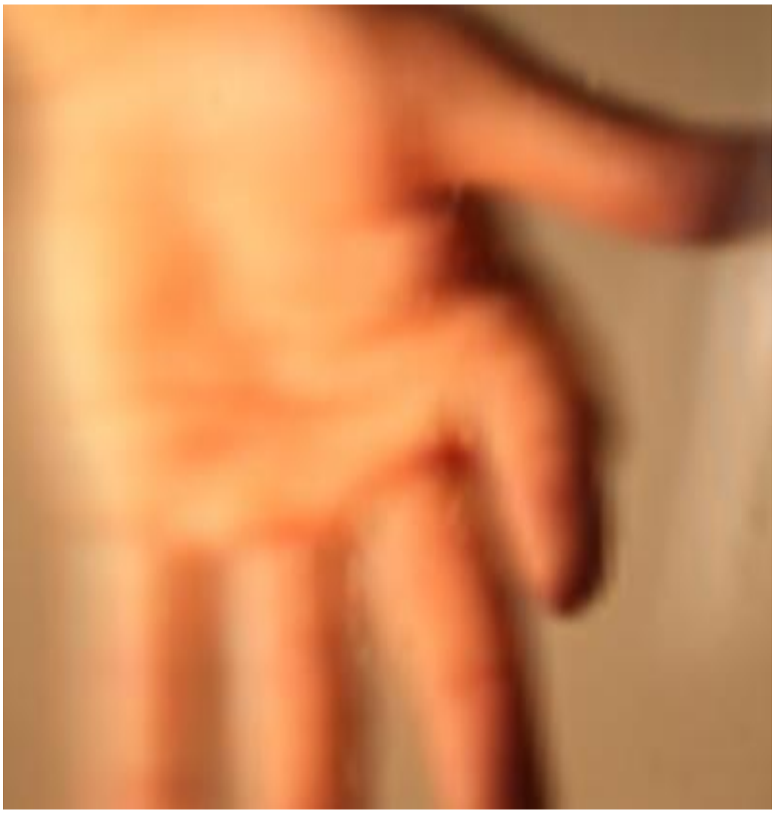
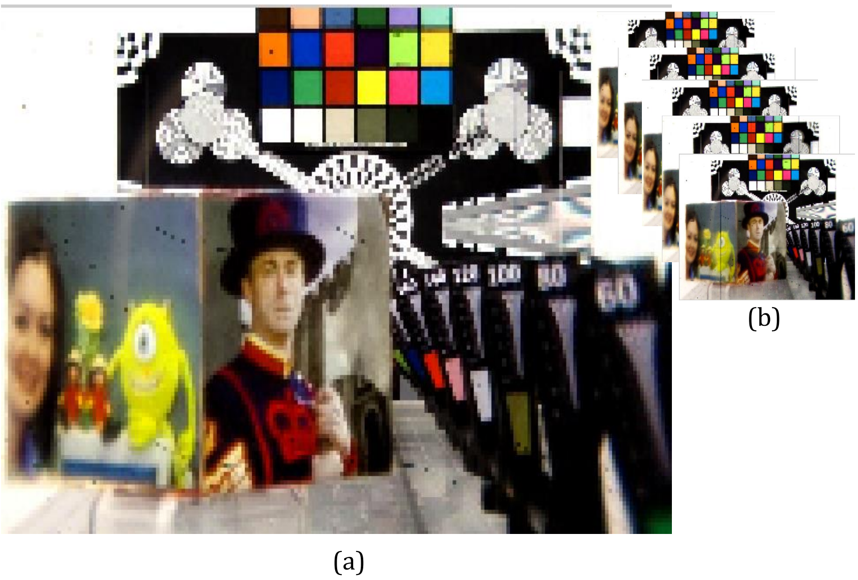


Figure 5-6 Result of the proposed algorithm on "Palm" real-world OH3DI using 18 sets (N = 18) of LRVPIs shows clearly the over smoothing SRVPI.

There were also significant improvements over the images in (c) and (d) that were obtained using the frequency domain algorithms and spatial domain algorithms respectively. However, (e) used the bi-cubic interpolation and this is outperformed by the results from (c) and (d) with methods that can be used to obtain 2D signals from many sets of images, but they are computationally complex, which is their limitation. This means they can really only be used in domains where real-time reconstruction is not needed. This shows that the new method produces images with sharper edges and less noise as well false edge irregularities when compared to the other algorithms. More details for the evaluation are available in Chapter 6 section 6.5.2.

5.3.2 OH3DI Experimental Results

The SR technique was applied to full parallax real-world OH3DIs to test the proposed approaches' performance. The "Box-Tags" and "Airplane-Man" images, both are obtained at 5616 × 3744 pixels, consisting of 193 × 129 micro-lenses, with EI resolutions of 29 × 29 pixels. The extracted VPIs from the first stage of the proposed algorithm had the same resolutions as the number of micro-lenses (193 × 129 pixels). Figure 5.7 shows the sampling stage of the "Box-Tags" VPIs. The sample rate integer factor was five as the SRVPI was created from fusing five sets of LRVPIs with each VPI shifted in the horizontal and vertical direction by one sub-pixel before being integrated in the second stage. A kernel size = [15 15] and (σ) = 2 was used in the post-processing stage for de-blurring of the SRVPI, which is demonstrated in Figure 5.8 for the "Box-Tags" image. Figure 5.9(b) shows the blurred SRVPI obtained after implementation of the process of up-sampled-shift and integration, while a magnified region is shown in (a). The image de-blurring filter is used to remove the blurry noise effect in the SRVPI in a post-processing stage that results in sharper edges, as shown in Figures 5.10 (a, b). The overall sharpness is not affected but the steepness of transition between the edges is enhanced to make them more prominent. This gives a more natural blurring effect as the region that is in-focus has the blurring noise removed while the defocused region is smoothed, as shown in Figure 5.9. As can be seen from the magnified region (a) of the de-blurred SRVPI the display is much improved and the edge can be seen in the final image (b) compared to the results in Figures 5.9(a, b). Figure 5.11 confirms the effectiveness of the new method by using the "Airplane-Man" image that has objects at a range of distances from the camera.



(b)

(a)

Figure 5-7 Sampling stage of the proposed algorithm on "Box-tags" real-world OH3DI, (a) one of the five LRVPI size 193 × 129 pixels used as reference for up-sampled shift and integration algorithm and (b) a set of the 5 LRVPIs been used in the experiment.

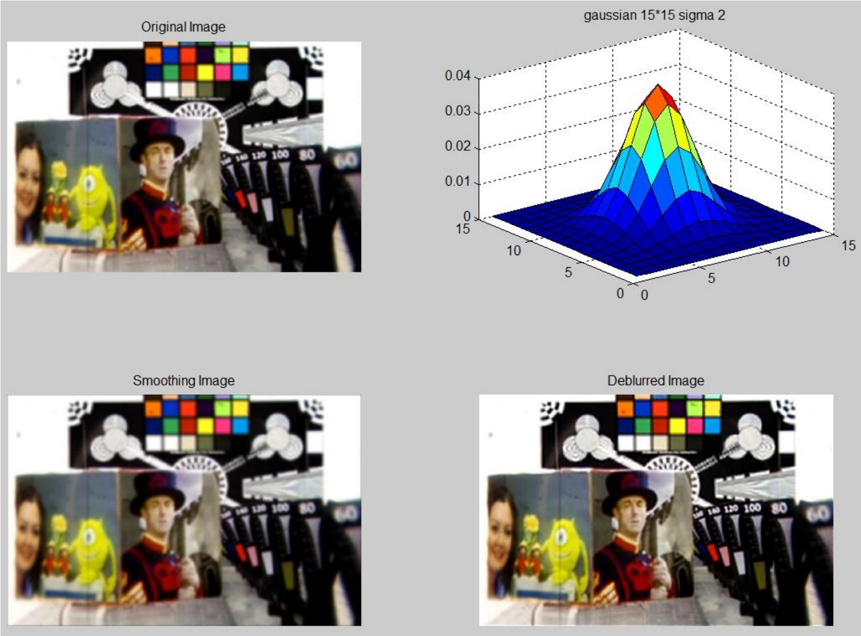
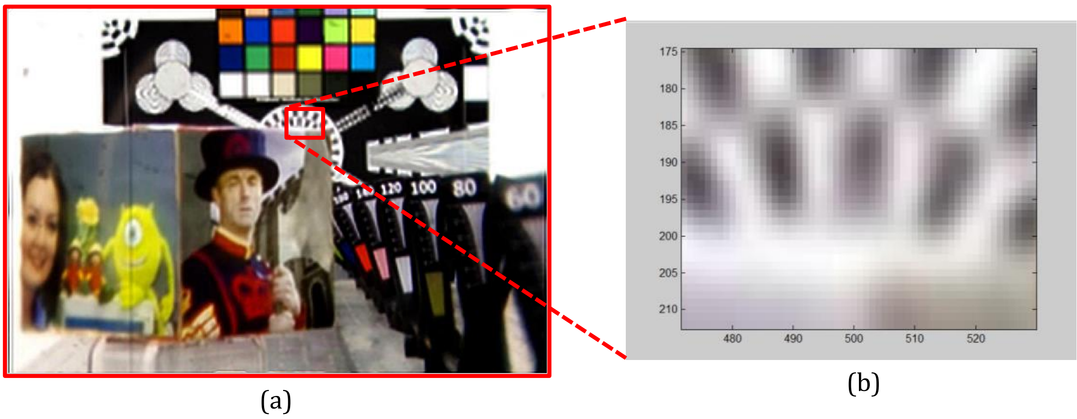


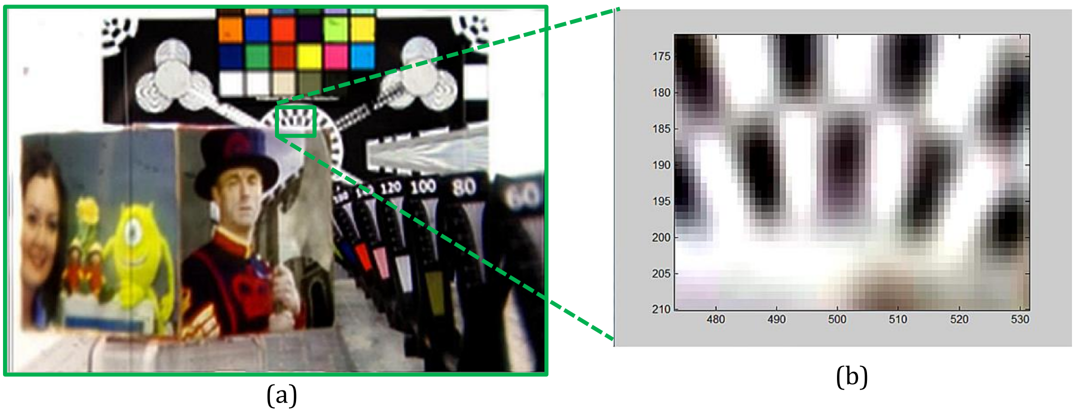
Figure 5-8 Post-processing stage of the proposed algorithm on real-world OH3DI, de-blurring process on “Box-Tags”.



(a)

(b)

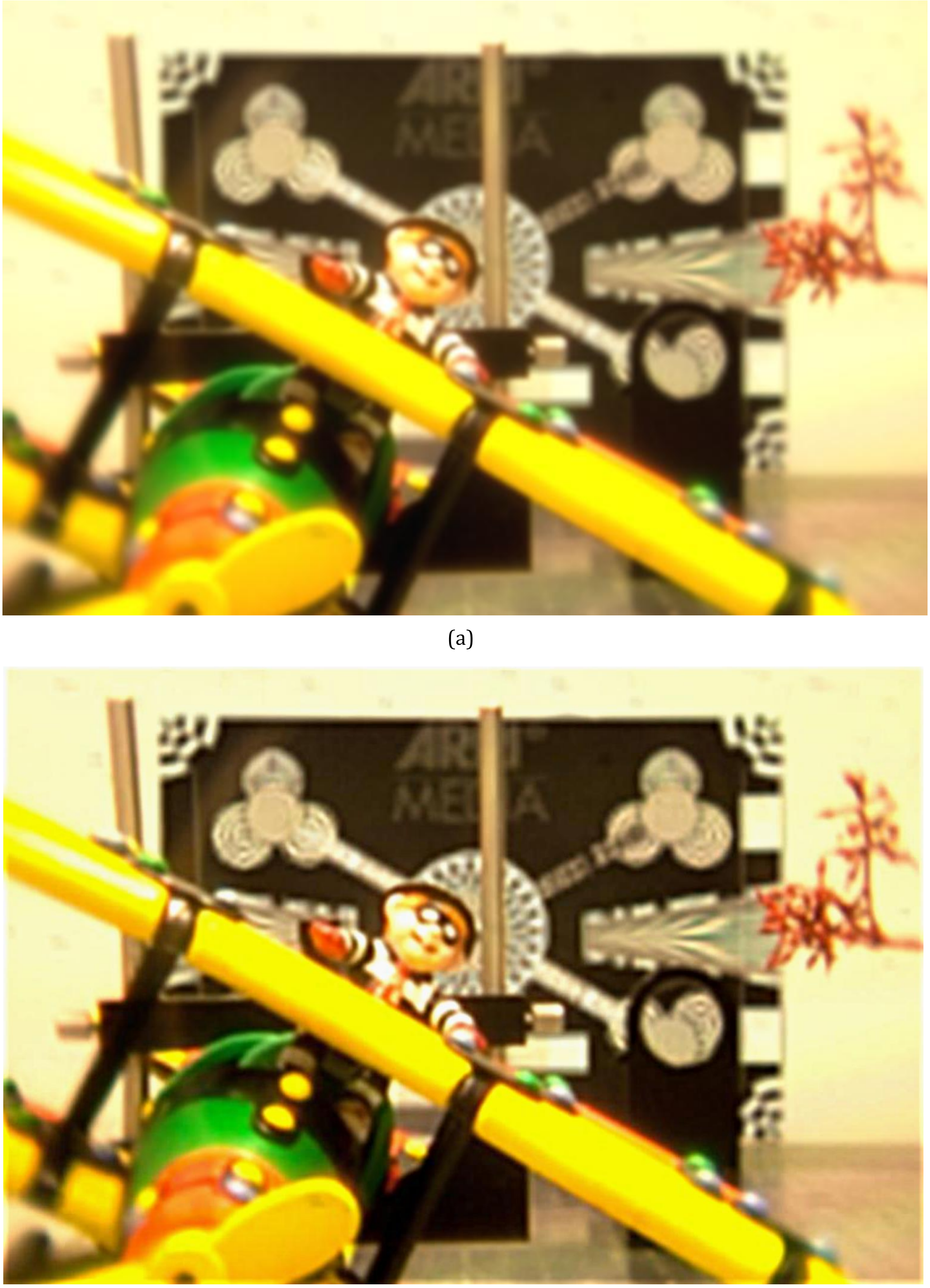
Figure 5-9 Results of stage up-sampled, sub-pixel shift and integration of the proposed algorithm on “Box-tags” real-world OH3DI, (a) a magnified region of the blurred SRVPI in (b).



(a)

(b)

Figure 5-10 Results of final stage (post-processing) of the proposed algorithm on “Box-tags” real-world OH3DI, (a) a magnified region of the de-blurred SRVPI in (b). The size of the de-blurred SRVPI is 973 × 653 pixels in perspective projection geometry.



(a)

(b)

Figure 5-11 Results of real-world “Ariplane-man”OH3DI, (a) blurred “Airplane-Man” SRVPI and (b) de-blurred “Airplane-Man” SRVPI.

5.4 Generating Multi-view Images from Holoscopic 3D Imageing

The up-sampling, shift and integration approach as well as the de-blurring smoothing filter are used in this section to render different perception views from a H3DI system. However, to use a H3DI system to produce multi-view SR images is difficult as the size of the image sensor limits the use of wide viewing angle micro-lenses. When the proposed SR algorithm is used to generate multi-view SRVPIs it is not necessary to have large baselines in the viewing. In general, five to nine perspective projection views are used in the multi-view auto-stereoscopic techniques [179], while the VPIs from the H3DI technique are orthographic projection LR images. As previously stated in Chapter 2, the capture process in the 3D multiple views system is expensive due to multi-cameras as well as the calibration process is linked to their use. The use of multi-cameras is also not suitable for content producers as they are not very mobile, when compared to the H3DI system. Therefore, to produce seven views without complex and expensive multi-camera calibration the holoscopic capture process is suggested.

It is possible to generate a composite VPI that is formed from the same scene as recorded from different perspectives through the use of a H3DI pixel format. This can be converted into a multi-view 3D image pixel format with the correct slanting. The following explains the rendering process where seven views are obtained from each EI of resolution 29 × 29 pixels. The process to generate the SRVPIs from the VPI's is shown in Figure 5.12 in which different colours (1-29) correspond to the EIs' positions. As also shown by different colours (SR1-SR7), each SRVPI view is maintained at a constant distance from each other in the rendering. For example, each EI has five views extracted from it to generate SR1. Within the EI, the first view is at x = 5 and y = 5, while for the second view is at x = 8 and y = 5. The y-axis is kept constant while the x-axis position is changed so that it is different but the distance between them remains the same otherwise there will be misalignments in the projection of the views causing bad 3D affects when displayed. Then the SR1 is rendered using the same parameters as in the previous section, with the integration algorithm to implement the up-sampling-shift. The horizontal direction is shifted by three views: that is, between pixel 1 in view 1 and the same pixel in view 2 is a difference of three pixels, and this is done for all seven views. As well as for the seven extracted views the parameters are set to be constant throughout. As shown in Figure 5.12, each view has a difference of three pixels for each view's axis position under each EI. Figure 5.13 shows the seven SR views that have been produced from the H3DI system.

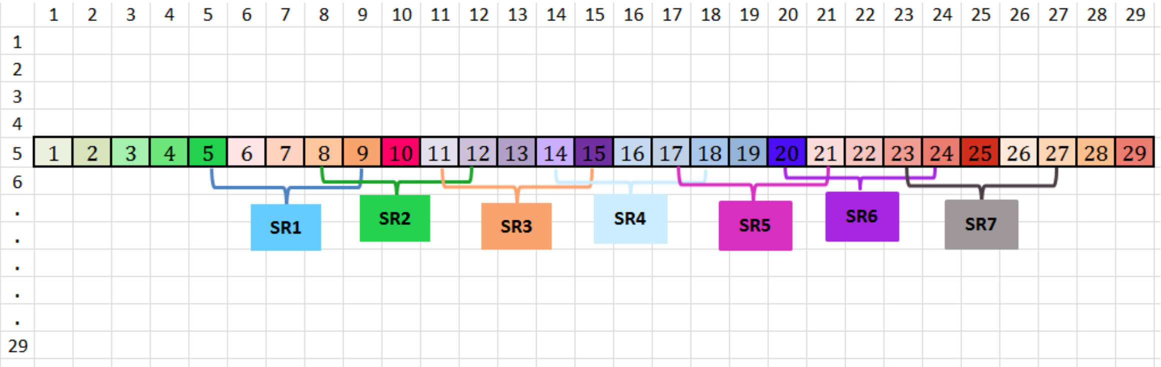
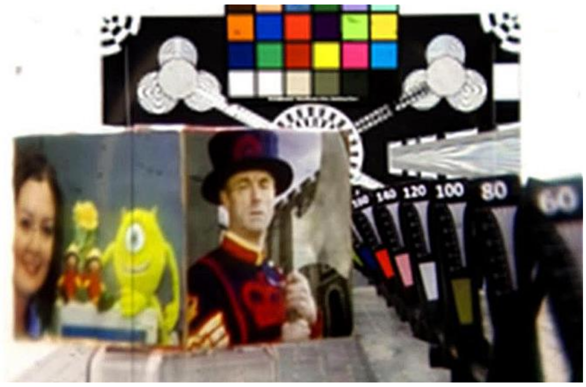


Figure 5-12 VPIs selection process to generate 7 SRVPIs.

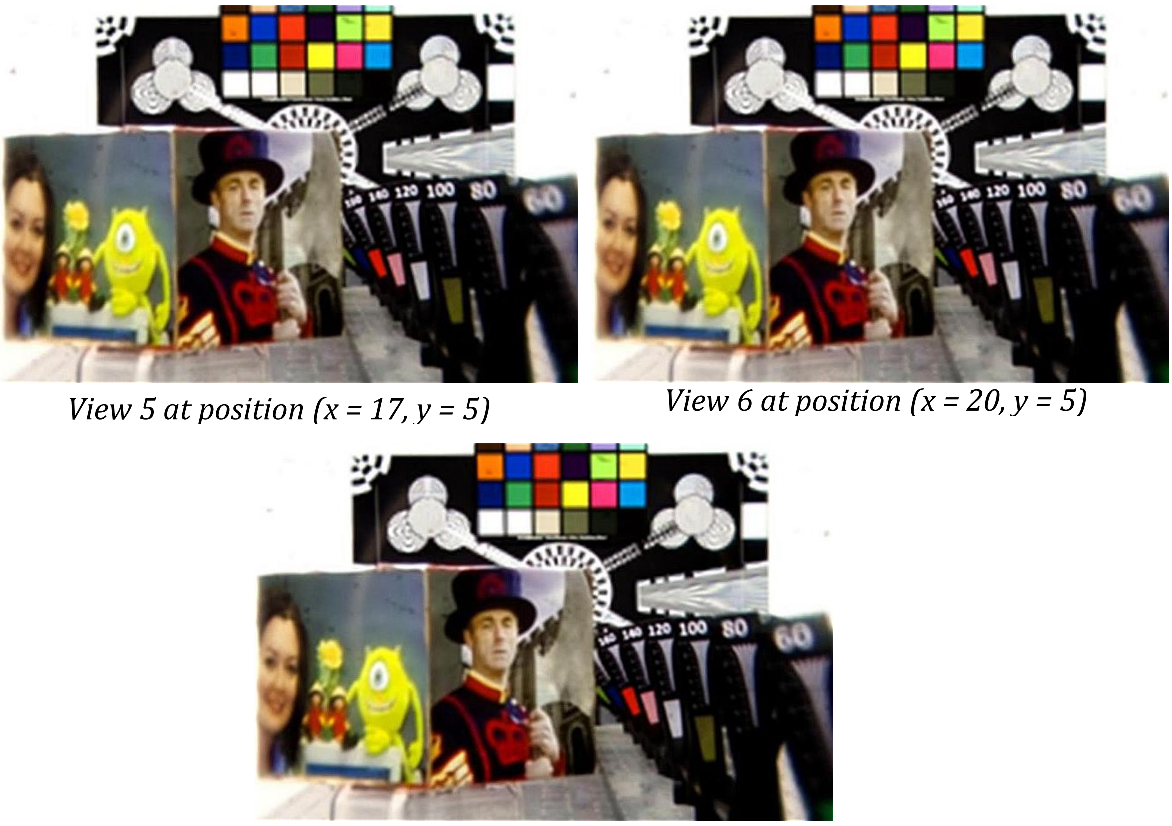


View 1 at position (x = 5, y = 5)

View 2 at position (x = 8, y = 5)

View 3 at position (x = 11, y = 5)

View 4 at position (x = 14, y = 5)



View 5 at position (x = 17, y = 5)

View 6 at position (x = 20, y = 5)

View 7 at position (x = 23, y = 5)

Figure 5-13 Display of 7 “Box-tags” SRVPIs generated from a single H3DI.

5.5 Depth Map Estimation using Multi-view SRVPIs

Due to the size of the camera sensor the VPIs extracted from the H3DI system have limited resolution, as stated previously. In this there is a trade-off between the baseline and the views. This means that the resolution and depth perception of the views generated from larger image sensors are greater. When using a larger image sensor with the multi-view auto-stereoscopic 3D imaging technique the views will be in perspective projection rather than the orthographic views from the H3DI system. Techniques like these are very expensive to use when capturing the image as they need multiple camera configurations. In addition, the brain is required to join separate images to make the 3D perception. This can cause eyestrain and headaches for the viewer due to unnatural viewing caused by focusing on the screen plane and the convergence of their eyes on a location in space at the same time [180]. Thus, the 3D technology is enhanced when the advantages from both techniques are combined.

Both depth accuracy and a fast execution time have been achieved to improve the depth map estimation field obtained from the H3DI system via the use of the multi-view SRVPI's characteristics. The estimation of the depth map from the holoscopic technique is achieved directly from increasing the baseline between the views and the resolution of the VPIs. This is directly related to increases in the robustness of candidate feature edges/points during the setting and extracting of the feature blocks process. There is also a reduced complexity in the computation when the H3DI system is used to produce the 3D image from the VPIs in the depth estimation. This method nullifies the two main problems linked with the holoscopic system, which are the VPI's limited resolution and computational complexity, by combining the advantages from multi-algorithms to optimise the performance implications in the real-world.

This approach is very clearly focused as a simple and efficient way of generating a high resolution depth map from seven sets of multi-view SRVPIs with a long baseline, which improves the disparity estimation's accuracy. Three techniques are combined to achieve this: 1) generation of SRVPIs by reformatting the extracted VPIs from orthographic (i.e. low resolution) to perspective (i.e. high resolution) projection geometry (see previous section). This sets reliable feature information blocks that are vital for improving the feature matching algorithm. 2) AFE detection algorithm for setting and extracting reliable 3D information (see Chapter 4), which is the key to success in realizing reliable features on SRVPIs for the next stage. 3) An adaptive hybrid multi-baseline algorithm (see Chapter 3) to improve the performance of the depth estimation and simultaneously maintain a low computation time. Figure 5.14 shows the performance of the multi-view SRVPIs algorithm for depth map estimation and the proposed algorithm's encompassing framework.

When the multi-view SRVPIs algorithm is used for depth map estimation the principle steps are:

1. A set of EIs are selected for the process.
2. Using the strong interval correlations that are present in pixels moved by one micro-lens, transform the EIs into VPIs.

3. Use the new interpolation, up-sampling-shift and integration algorithm to form higher spatially dimensioned VPIs from the LR orthographic projection VPIs that have been converted into a perspective projection.
4. To enable the extraction of a set of reliable features select the reference VPI as a training image using the AFE thresholding algorithm (as in Chapter 4).
5. Employ the adaptive window shape (AWS) in the new version of the adaptive multi-baseline disparity algorithm that was highlighted in Chapter 3.

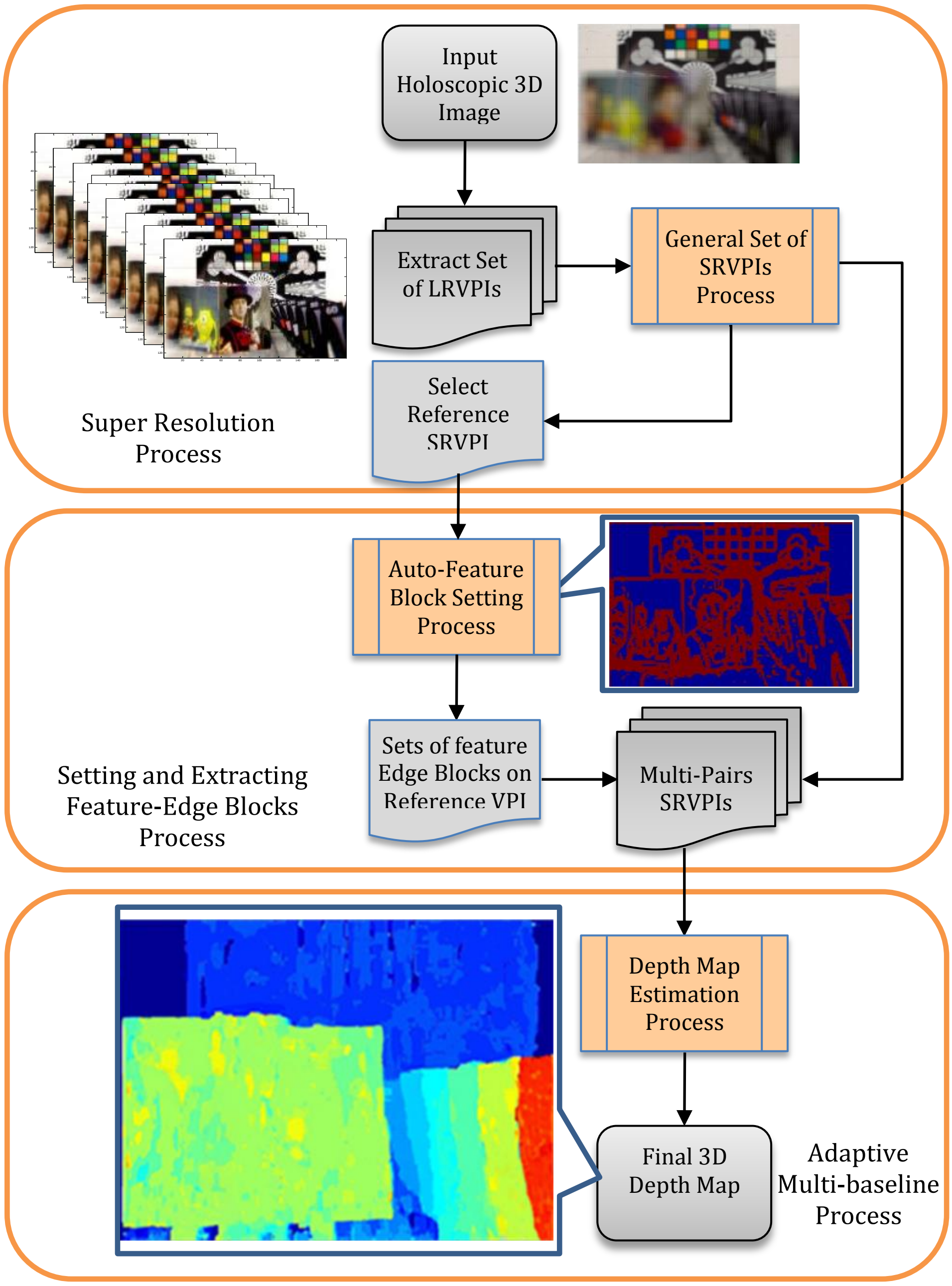
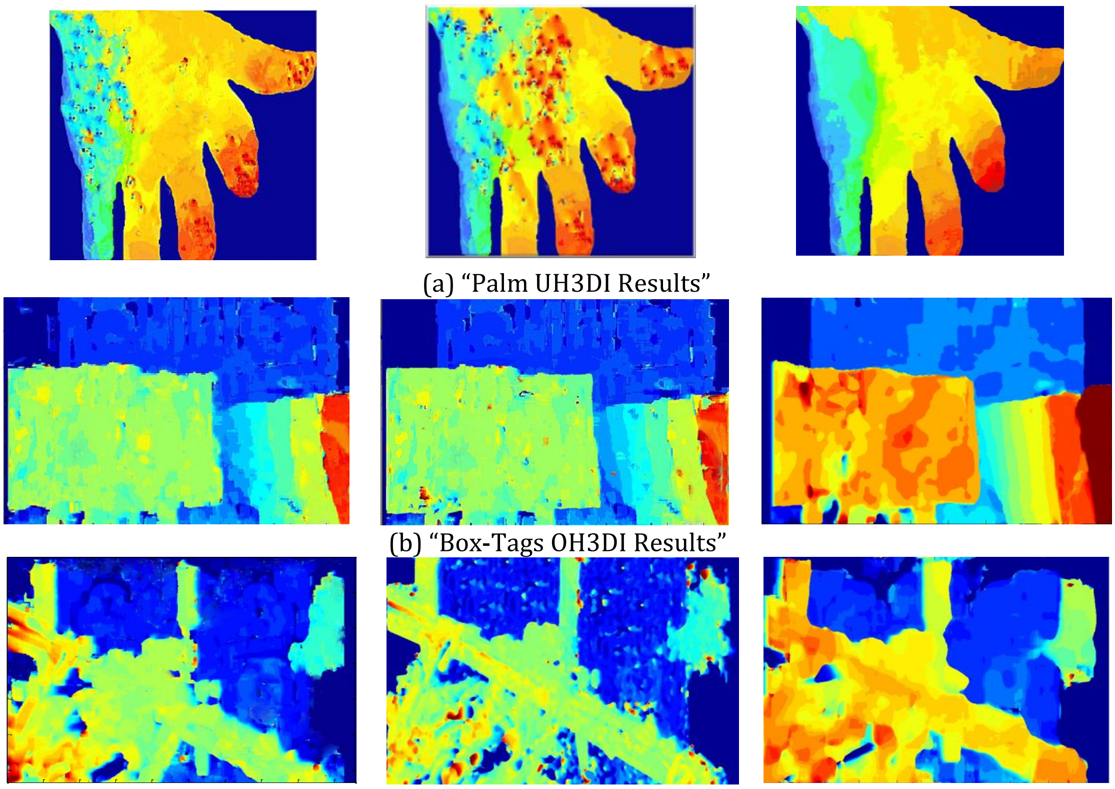

Input Holoscopic 3D Image
Extract Set of LRVPIs
General Set of SRVPIs Process
Select Reference SRVPI
Super Resolution Process
Auto-Feature Block Setting Process
Sets of feature Edge Blocks on Reference VPI
Multi-Pairs SRVPIs
Setting and Extracting Feature-Edge Blocks Process
Depth Map Estimation Process
Final 3D Depth Map
Adaptive Multi-baseline Process

Figure 5-14 Overall representation of the proposed procedure process for 3D high-resolution depth map estimation on real data “Box-Tags” OH3DI.

When compared to the other algorithms presented in Chapter 4, these experiments showed that the new method outperformed them when identifying the precise depth map results and the specific objects in a scene with regards to the accuracy and speed. In these tests the previous algorithms used sets of 49 LRVPIs to form the depth maps. Experiments have shown an obvious improvement on the depth map results achieved by the proposed method on several captured holoscopic 3D contents.

Figure 5.15 shows a summary of the depth map estimation results generated via a range of different algorithms. The results of the auto-feature-point (AFP) algorithm are shown in the first column, the second has those of the auto-feature-edge (AFED) detection algorithm and the third has those of the proposed high-depth estimation method. The third column clearly proves the improved performance of the proposed high resolution method from a set of SRVPIs. The rows in the figure correspond to the different H3DIs used with the first being the “Palm”, the second being the “Box-Tags” and the third being the “Airplane-Man”.



(a) “Palm UH3DI Results”

(b) “Box-Tags OH3DI Results”

(c) “Airplane-Man OH3DI Results”

Figure 5-15 Comparison of the quality of depth map estimation results using various images with different algorithms.

5.6 Summary

This chapter explained the depth map creation by reformatting holoscopic VPIs to perspective view image and how a H3DI technique can be used to generate multi-view SRVPIs with computational SR method. Within the SRVPIs there is ample information that can be used for depth estimation. This method produces a single view direction of the recorded scene by registering a set of VPIs. This registration is based on forming the VPIs by sampling the same pixel in the same location under different micro-lenses. This means that the set of LRVPIs have already been registered. The reconstructed SRVPI is made by reformatting a set of sample VPIs so that it has known offsets within a multichannel sampling problem. The new method transforms and refocuses the set of LRVPIs orthographic projections into a SRVPI perspective projection. The visual quality is improved through the use of a new interpolation approach that adds high frequency information contained within the scene to the final VPI.

It has also been shown that when using the new SR algorithm it is possible to generate multi-view SRVPIs from a H3DI system. These multi-view SRVPIs have been used successfully for depth map estimation due to the rich information they contain as these results come from large image sensors. The information is successfully used for object detection and recovery, and then employed to extract robust correspondences, thus leading to a reliable estimation of 3D object depth. In this proposed approach the algorithm is implemented for the high resolution VPI to extract reliable feature information blocks by searching for the optimal threshold value, which guides the setting and then extraction of a reliable set of features.

This chapter presented work that was focused on developing algorithms to extract 3D depth maps by reformatting holoscopic VPIs to perspective view images in the simplest way. It was shown that to determine the correct object from the scene with high accuracy the disparity analysis was very important. Exhaustive experiments demonstrated the consistency and efficiency of the proposed method to extract 3D object boundaries surfaces and forecast mislaid depth values for texture-less regions. The proposed method outperforms existing algorithms for overcoming 3D measurement based object boundary limitations using a MATLAB simulation environment.

CHAPTER 6

Holoscopic 3D Image Segmentation

6.1 Introduction

3D imaging systems are judged to be successful when a view synthesis of suitable quality is obtained at the receiver. Generating a depth map with high quality synthesis is a challenging task when there is a lot of noise in the image and there are no extra cues related to the scene available. This is often due to passive pre-pixel depth estimation in the 3D systems. The methods used in computer vision are based on stereo principles rather than utilizing laser range scanning or computer graphics models. There is also noise introduced at the receiving end due to the quantization of the pixel depth [181]. This chapter presents a segmentation method for a H3DI system that offers interactive depth mapping of integral scenes.

3D image segmentation systems are used to segment and determine the objects presence in a scene as the objects needs to be identified despite the noise, partial occlusion and clutter in the scene. The approaches utilizing 2D cameras produce a fast result when there are no changes in illumination, shadows or occlusions as these cause them to fail to generate an accurate estimation of the object's pose [182]. Researchers have proposed good potential algorithms for image segmentation that are based on simple edge detection and more domain specific algorithms such as with eye tracking algorithms, which have been used in a wide range of application areas. Despite being able to utilise a large amount of previous research when developing new computer vision applications, the robustness to cope with the effects of various noise is still lacking.

“Segmentation” refers to an image being partitioning into various sections in computer vision. Nevertheless, the proposed work focuses on separating the foreground from the background. This work does not just focus on segmentation algorithms despite them being the end objective. In addition, the interactive algorithm will be used to generate the base-map that is used to determine the regions of interest in the image. Thus the basis map will be referred to as an interactive map. This map provides the information about the areas that will be of interest to the viewer, which should be the focus. As far as this work is concerned the objects in the scene are defined by the object segmentation and by the 3D-Interactive-Map (3DIM). The proposed 3D object segmentation technique is a hybrid method that combines an existing 2D segmentation technique with the 3D depth cues. This means that the 3DIM uses information obtained from the 2DVPI segmentation module and the 3D depth estimation map. This enables the advanced object registration that is needed by editing tools. The proposed 3D object segmentation framework is shown in Figure 6.1. The visual cues and interactive map are encoded together, so that additional cues about the contents can be correlated.

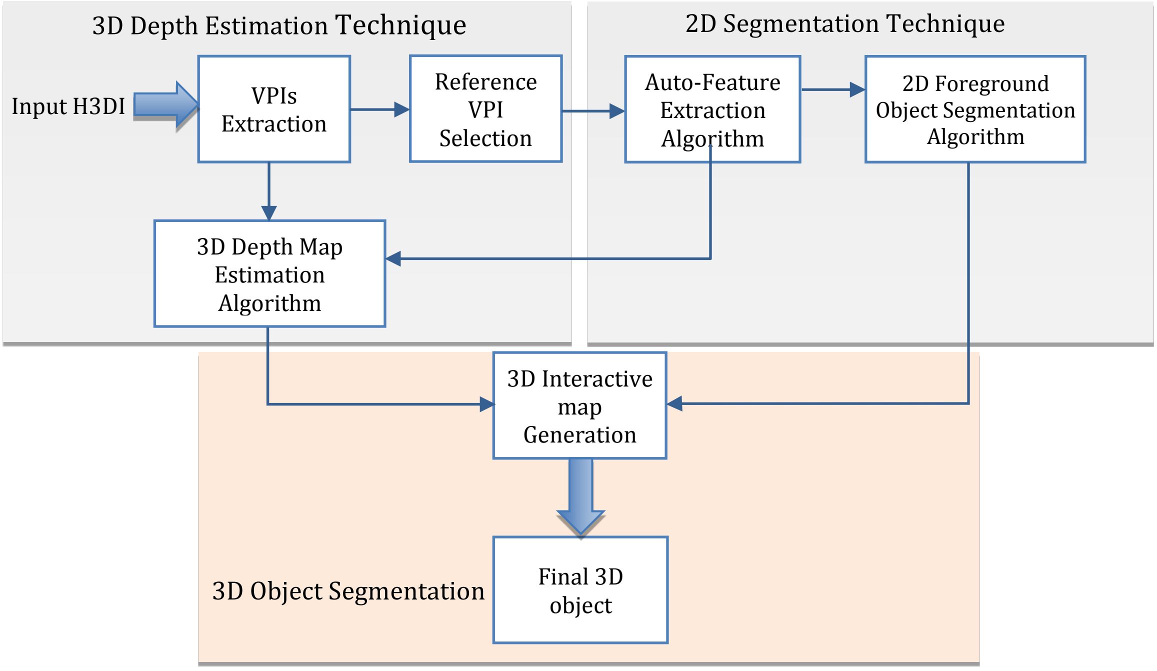

3D Depth Estimation Technique
2D Segmentation Technique
Input H3DI
VPIs Extraction
Reference VPI Selection
Auto-Feature Extraction Algorithm
2D Foreground Object Segmentation Algorithm
3D Depth Map Estimation Algorithm
3D Interactive map Generation
3D Object Segmentation
Final 3D object

Figure 6-1 Block diagram of the 3D object segmentation framework.

This means that for a 3D object to be recognized the related foreground regions and depth information map needs to be identified. To be able to use the H3DI system to produce the 3DIM model, the segmentation of the extracted 2DVPI is essential. The module gives cue of the object in the scene via incorporating the shape before the background noise/objects are removed in the resulting depth map. In other words, the segmented VPI is active as a segmentation filter mask module. As far as this work is concerned, the most important filters are the feature extraction filters, which produce binary mask feature maps that are represented as "0" or "1" values. Zeroes denote non-salient regions, while ones denote salient regions. This mask map is combined with the depth map's output and gives the descriptive output (as shown in Figure 6.1). For 3D image segmentation from H3DI systems, there is no solution despite there being many general segmentation, methods. It was necessary to make yet another segmentation, as part of this work because all the current ones use a set of fixed algorithms that are combined in a fixed configuration. This means that using an adaptable environment over a fixed one is a way that successful data input processes can be identified.

This chapter focuses on the separation of tangible 3D objects from a 3D scene using a 3D image segmentation algorithm. It has always been a great challenge to effectively extract the foreground surrounded by backgrounds to distinguish the object in the scene from the background in a single image. For a 3D object to be recognized the related foreground regions and depth information map needs to be identified. This was achieved despite the area of research being new. Two novel unsupervised segmentation methods that generate interactive depth maps from single VPI segmentation were suggested and tested on real data and C.G. H3DIs. Both techniques offer new improvements over the existing methods due to their simple use and being fully automatic; thereby, producing the 3DIM without human interaction. Section 6.3 and 6.4 give discussion, implementation and results for each method.

6.2 Related works

There are two types of H3DI reconstruction techniques, namely: optical and computational [183, 184]. Based on ray optics, a lenslet array is used in the computational H3DI reconstruction technique to inversely map EIs. The object images are reformed as depth plane object images in the output plane, where the 3D object is originally located [185]. This means that the depth data of 3D objects in a scene can be extracted when the foreground (object) can be separated from the background (noise or any other objects on a different plane).

The capability to reconstruct the depth plane of 3D objects along the output plane is a unique feature of the computational H3DI reconstruction technique. Using the estimation of the blur metric of each image from the object plane, the computational reconstruction of depth plane data from a H3DI technique has recently been accomplished to generate a 3D object [185]. When the points of inflection are examined it is possible to use the blur metric to see which objects are or are not in focus. The blur metric is low when the image is in focus and high when the image is out of focus. This enables the object's original output plane to be identified and the depth data extracted. This has been proven by experimental results. However, near the edges of the objects the performance becomes worse as there are many local maximum points in the contour plane for instance when there is a small change in the values of the blur measure a point might be between two planes. This method is tested using objects that are positioned only 60 mm in front of the camera (Figure 6.2). This means that when used with objects in the real-world that are much further away there can be problems with the small discrimination values of the points of inflection that occur in the middle of the located object points' plane and neighbour planes.

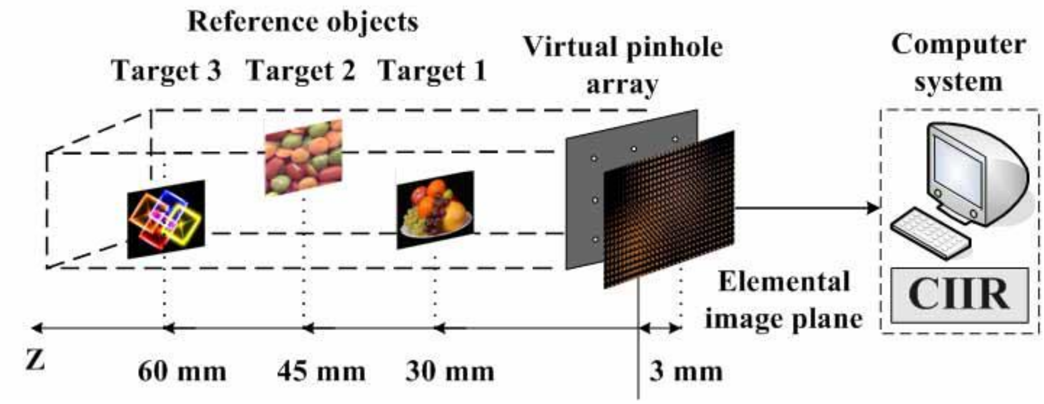


Figure 6-2 Experimental H3DI system setup that uses pinhole array [185].

At the moment, the use of the segmentation technique on a VPI has been presented [186] and it uses a novel feature approach for 3D object segmentation from a H3DI. This method uses a graph based approach in a multi-level system to integrate various analysis and processing algorithms to form an environment for use with any type of 3D object segmentation. The first step is to use a smooth filter from the segmentation technique on the 2DVPI to change the extracted VPI into the frequency domain so that the filter can be processed in a rapid way. It is assumed that for the spectral residual the average VPI has a smooth spectrum that enables the 3D object to be located in the scene; anything that is not smooth should be something that the viewer needs to recognize. Then a Canny edge detection algorithm is used to determine the feature edge from the original VPI for object separation, as well as the closing of small gaps that might be present around the edges of the objects via a morphological closing operation. Then all the values that are set to 1 are counted to determine the objects from the background (set to 0) to form a map that has been transformed into a logical map using the smooth spectrum filter map. Then the feature edge map is generated by starting from the maximum value that is found in the smooth spectrum map. The final result is a map where objects are shown by values that are "1" and the background by "0".

Experimental results have shown that 3D objects in the scene can be correctly identified by using the previously described approach by producing a separation 2D filter map from the H3DI technique. When optimising the result it was found that the VPI's size was a critical factor and it was possible to get adequate results when the input data had a size of 64 × 113 pixels. The better the resolution, the fewer details were effected. However, if an object is used that has a resolution that is too high then it will no longer be considered a complete object. A lower resolution generally gives a more defined outline for the objects, which is what the algorithm detects. This means that the algorithm is only suitable for use with specific sized holoscopic 3D content: it lacks generalisation. In addition, for the algorithm to function successfully the determined edge map must be closed and this depends on the quality of the calculated feature edge map. In this work the Canny edge detection algorithm has a range of values that can be adjusted, and these settings have an impact on the algorithm's effectiveness. When smaller Gaussian filters are used there is less blurring that enables smaller and sharper lines to be detected in the Canny algorithm. In contrast, if a large filter is used, then the values generated from each pixel will be spread over a larger area resulting in blurring. The threshold needs to be set to the best position in the middle since too low noise in the image is determined to be important then when it is set too high, objects will be missed. As often each image needs a unique threshold therefore it is not possible to set a generic threshold, which works on a wide range of images [187].

6.3 3DIM Generation Based on Intensity Segmentation

Generating a well segmented 3D object level requires an algorithm that is capable of identifying two items: foreground/background regions and their boundaries. This work focuses on the segmentation of foreground objects from background regions that are in the scene starting from the centre of the image. It focuses on improving accuracy of the foreground extraction that produces an accurate 3DIM. This provides an improved solution to the 3D segmentation problem by integrating 3D information techniques and 2D foreground object segmentation technique. The algorithm proposed in [188] is used to extract the foreground map using a morphological operation for the reconstruction of the 3D object.

Morphology consists of various operations for processing the image that use shapes, and there are many examples of morphological segmentation algorithms in the literature. While another group consists of mathematical morphology, within which is a set theoretic shape oriented method that considers the image as a set and the kernel operation is referred to as a structuring element (SE). This means that the input and output images are the same size when the SE is used: the output pixel values are based on the corresponding value from the input and those of the neighbouring pixels. The SE used and its size determines how many pixels are added or removed from the objects located within the image. A strong mathematical concept is the basis of the morphological operation and is used to alter the size, shape, structure and connectivity of the objects found within the image, and consists of a range of binary and grayscale operations.

6.3.1 Morphological Reconstruction

This algorithm used the following morphological operators:

1. *Erosion* is used to remove pixels from the edges of objects in the image. It splits joined objects or shrinks objects by removing extrusions. $f(x, y) \ominus SE$ gives the erosion of an image $f(x, y)$ by a *SE*. This means that all the pixels under $f(x, y)$ are in the foreground and should be left alone, but if not, then the current pixel is in the background and should be classified as such. The following formula can be used to identify the new pixel value where the *SE*'s original place was (x, y):

$$g\ (x, y) = \begin{cases} 1 & if\ SE\ fits\ f(x, y) \\ 0 & otherwis \end{cases} \tag{6.1}$$

2. *Reconstruction* is a method for identifying shapes in an image and providing useful cues related to them. It can extract marked objects, define regions of brightness that are surrounded by areas of dark pixels, find and remove, if necessary, objects that cross the images' edge, find and then fill object holes, remove areas that are abnormally high or low as well as many other operations [189]. The method is very powerful when used to identify objects that are completely within the image. The method needs two images and the *SE*. One image is the initial point that acts as a marker, while the other confines the transformation and acts as a mask. The process starts from the peaks in the marker image that is developed from the mask image. Then the peaks start to spread and fill the mask image until the values become constant. When the SE is located at the origin, then K and M are the mask and marker images, respectively. The following equation defines the reconstruction from M to K:

$$M_{new} = (M_{old} \oplus SE) \cap K \qquad where,\ M \subseteq K \tag{6.2}$$

3. *Dilation* inserts pixels around the object boundaries in the image. It enables breaks to be repaired or objects to be made bigger. It is denoted by $f(x, y) \oplus SE$ which shows the increase in size of the image $f(x, y)$ by *SE*. This means that when the current pixel is in the foreground, all the other pixels under $f(x, y)$ should be the foreground in the output. The following equation determines the new pixel value when the *SE* has its origin at (x, y):

$$g\ (x, y) = \begin{cases} 1 & if\ SE\ hits\ f(x, y) \\ 0 & otherwis \end{cases} \tag{6.3}$$

Computational costs in the system can be reduced when dilating $f(x, y)$ by SE and the result should also be dilated by SE in a process called decomposition of the SE.

6.3.2 The Proposed Technique

The proposed method provides fully automatic 3D image segmentation by identifying the objects in the 3D scene to those from the 2D image, and the algorithm is called morphology-based-threshold (MBT). The method is based on the spatial intensity distribution in the image, which forms the simple thresholding technique. In the method, each pixel is defined as either an object point or a background point by giving each pixel an intensity value T (threshold). All the pixels in the image are assigned a threshold value independently of their sounding pixels, which means the threshold value is constant throughout the image. The method classifies high-intensity pixels as interesting and those that have a low-intensity as not interesting. Then the image is used with a specific threshold value to determine the regions of interest by generating a binary mask. As shown in Figure 6.1, the proposed MBT algorithm is used to identify the foreground objects in the scene. Then a 3D object can be produced that is noise free and in which the foreground and background are separated from each other by using the 3D information map (generated previously) and the 2D foreground mask. MATLAB software is used to run the MBT algorithm following the flow diagram presented in Figure 6.3. As stated previously, the morphological image processing uses a strong mathematical concept that incorporates a range of binary and greyscale operations. The method uses similar pixel intensities as given by the radius in the training reference VPI, morphologically manipulated operations of erosion, reconstruction and dilation that give a single value for pixel intensity. All of these methods have been discussed in detail previously. The following steps form the proposed MBT method:

1. Based on the intensities' variation from black to white in the image, the RGB training VPI is transformed into a grey level image.
2. Then the image I has all its features extracted using multi-scale morphological operations. The object in the image is segmented using a dual multi-scale reconstruction operation. The first step is to use the SE of a specific size to segment the image by opening reconstruction morphological operations.

3. This is more effective than using standard opening as the object's overall shape is not altered while small blemishes are removed. Then for the specific situation, the opening process is erosion followed by dilation. As shown in Figure 6.3(a), an SE value of 30 gives excellent segmentation while, extracting all the objects in the scene.
4. In step 2 i.e. the complement result, the background and foreground swap positions, with the background becoming the foreground and vice versa (Figure 6.3(b)).
5. The background and foreground from step 3 are then indexed using a threshold that is equal to the maximum intensity value in the object [185].
6. The algorithm from [190] is used to fill holes via set dilations, complementation and intersections. Starting at point p inside the boundary, all the background points are filled with "dark" values, while the foreground points are filed with "light" values. The following morphological operation is used to fill the "dark" areas:

   $$X_k = (X_{k-1} \oplus SE) \cap A^c \qquad (6.4)$$

   where, the initial assigned point is $X_0 = p$, the symmetric structuring element is SE and the previous step's output for set A is A^c. When the iteration step k is $X_k = X_{k-1}$ the algorithm terminates, where, as shown in Figure 6.3(c-d), the set union of (X_k and A) contains the filled set and its boundary. The region filling process is shown in greater detail in Figure 6.4.
7. Then to generate the 3DIM of the scene, the result from step 5 should be masked to the depth map (Figure 6.3(e)). The interactive content is created from the 3DIM that comes from the combining of the 2DVPI segmentation module and the 3D map to identify the objects that compose the scene. This basically means that the areas in the image from the objects and background are identified. Any regions that have been miscalculated will be dealt with and removed during the process.

The process described above can accurately separate the objects from the background and identify areas that have been miscalculated before removing them. When foreground mask map and the object are matched approximately it is possible for the filter to separate an object with a very detailed border from the scene.

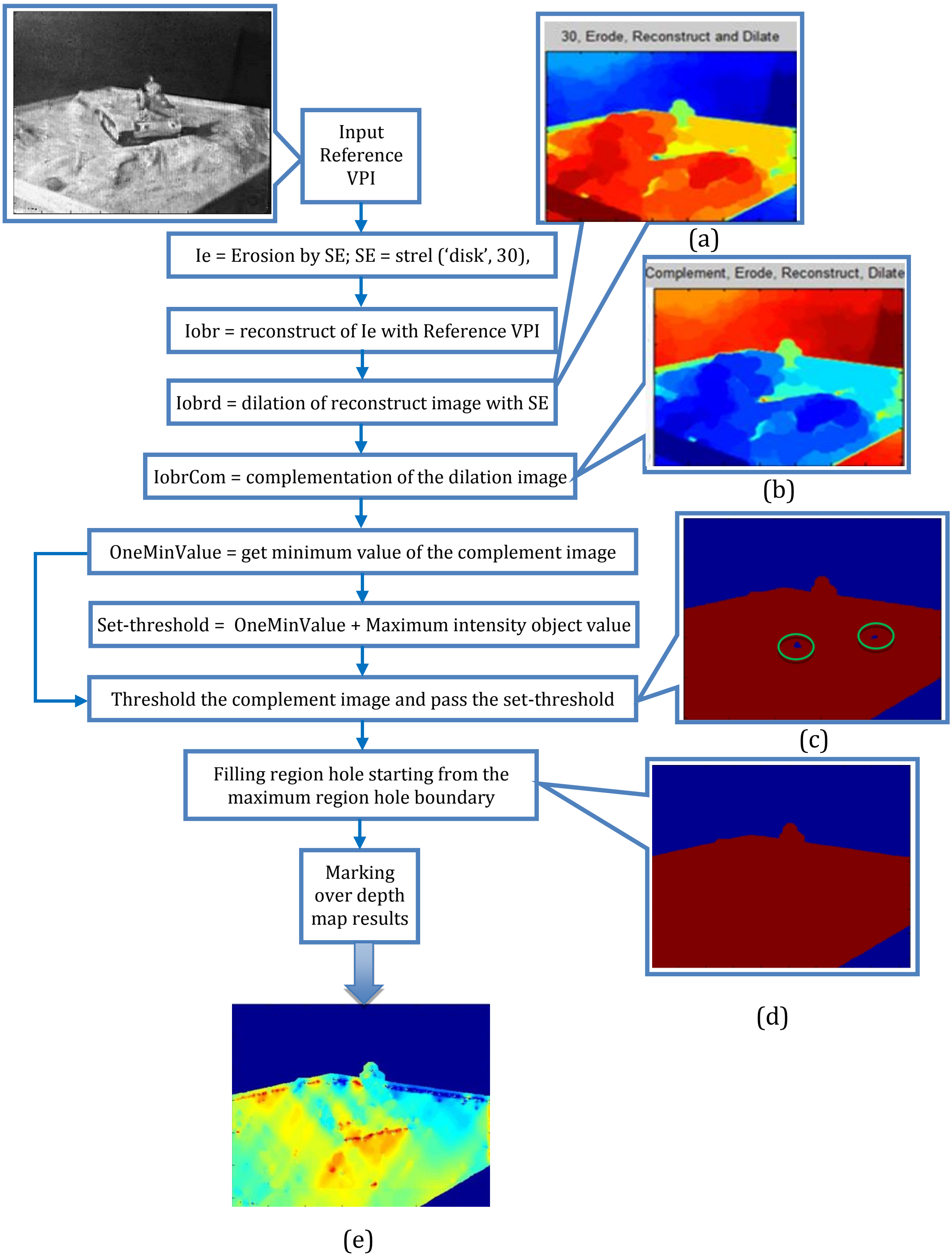

Input Reference VPI
Ie = Erosion by SE; SE = strel ('disk', 30),
Iobr = reconstruct of Ie with Reference VPI
Iobrd = dilation of reconstruct image with SE
IobrCom = complementation of the dilation image
OneMinValue = get minimum value of the complement image
Set-threshold = OneMinValue + Maximum intensity object value
Threshold the complement image and pass the set-threshold
Filling region hole starting from the maximum region hole boundary
Marking over depth map results
30, Erode, Reconstruct and Dilate
(a)
Complement, Erode, Reconstruct, Dilate
(b)
(c)
(d)
(e)

Figure 6-3 Proposed MBT algorithm flowchart and the results obtained from each step on "Tank".

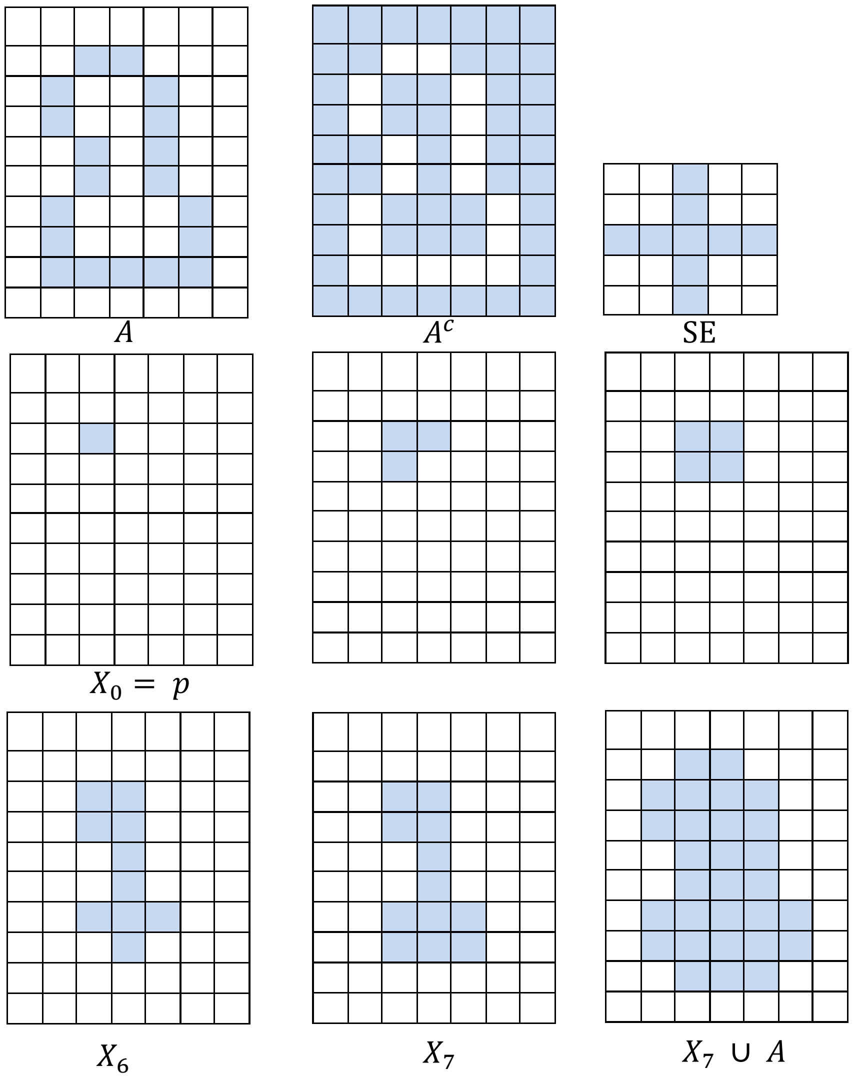
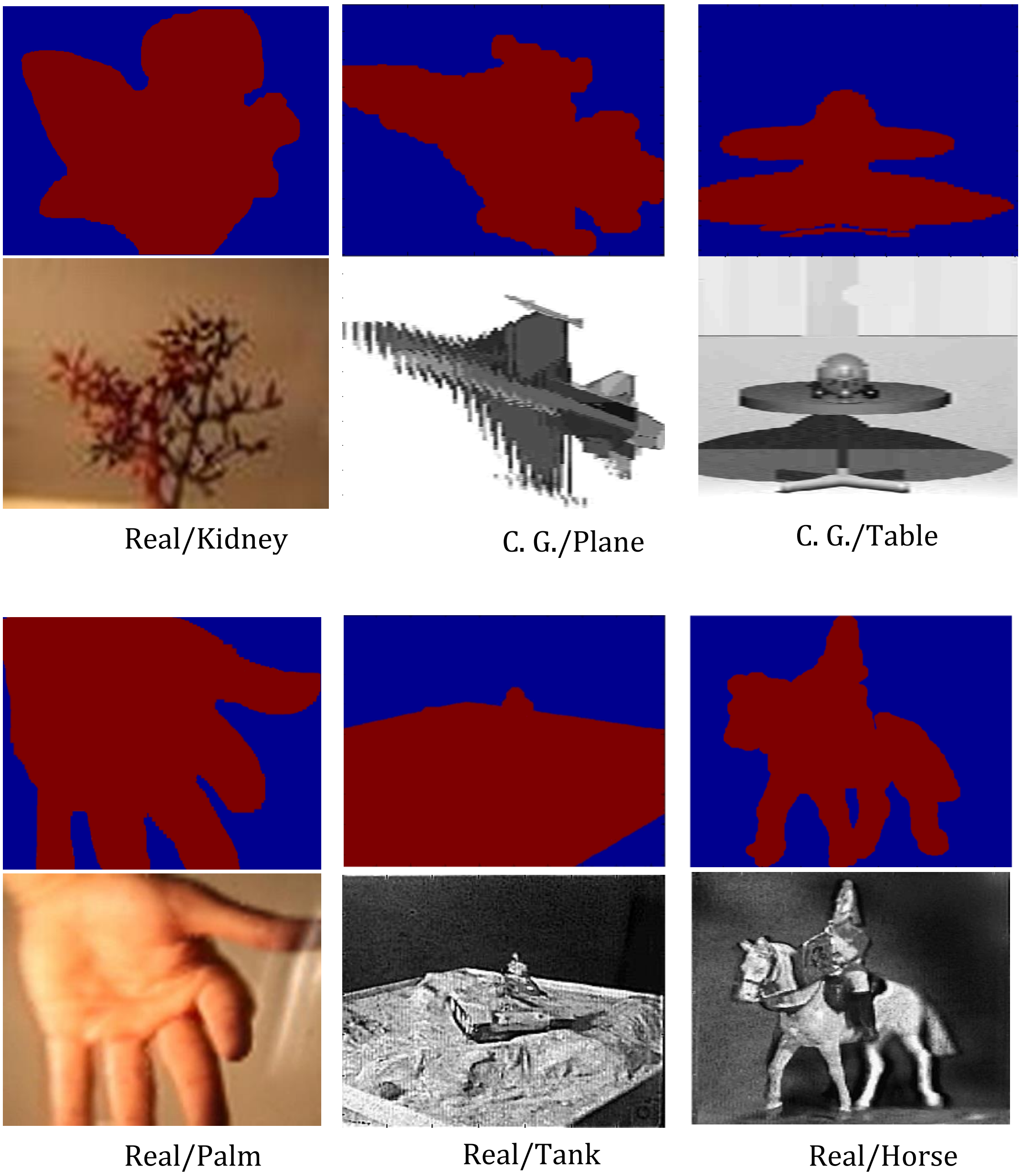


Figure 6-4 Morphological region filling process, (a) set A, (b) Complement of A, (c) structure elements, (d) initial point inside the boundary, (e-h) various steps of Eq. (6.3), (i) final result, which is the union of (a) and (h) [190].

6.3.3 Experimental Results

Several real-world and synthetic H3DIs with various texture qualities were used to determine the proposed MBT algorithm's performance. This section presents some of the results obtained, using the new 3D object segmentation method. The intended aim for the algorithm is to generate segmentation of the H3DI system during the process of producing a 3D interactive map. Two techniques are integrated in the 3D segmentation module, the first being the proposed 2D segmentation technique that identifies objects in the training VPI with a 3D information map estimation technique presented in Chapter 4. Some results are presented in Figure 6.5 for the real-world H3DIs (kidney, tank, horseman and palm) as well as synthetic images (plane and table). The results show that the regions of greatest interest from the images are represented by the 2D foreground masks. In the final output the foreground and background areas are clearly separated using the final output that is a binary mask. The background is a dark blue colour, whiles the foreground, as represented by the binary mask, is red. Then it is possible to generate a wealth contour map of the objects when using the foreground mask, which shows that it is suitable for object separation: it can determine the borders of the objects and remove them from the background. Once the map has been generated from the previous steps it can be used to improve object separation from the depth information map due to the borders of the objects being very pronounced or it can assist in the removal of noise associated with the images when a filter was used. Some of the 3DIMs that were used to test the framework of the 3D object segmentation approach (Figure 6.1) are shown in Figure 6.5. The algorithm works well as it prevents the background from being selected when determining the 3D object in the scene. In addition, the results showed that the 3D object segmentation approach relies on the quality of the foreground mask; while, the degree of match approximation between the filter masks and object is important when the object to be separated has a very detailed border.

Experimental results show that the contours closely belonging to the shape of an object and thus, used to improve the separation of the objects' depth information map, can also be implemented as a filter to remove noise associated with images. Accordingly, the 3D objects were correctly localized**.** However, section 6.5 includes both subjective and objective evaluations of the approach, which has identified some over-segmentation errors. This mainly occurred due to the acquisition of false edges of markers, which highly depends on the gray level distribution (high variation between background and object).



Real/Kidney

C. G./Plane

C. G./Table

Real/Palm

Real/Tank

Real/Horse

Figure 6-5 Outline of the proposed MBT segmentation algorithm, representing the foreground mask results. The first and third rows display the foreground mask for real-world and C.G. H3DIs data and the others rows display their training VPIs.

6.4 3DIM Generation-based Feature-Edge Extraction

This section details the integration of the result of the edge detection algorithm obtained by the local histogram analysis process to form a new segmentation technique for the H3DI system. This method is used to identify the objects and separates them from the background. This combines a technique of grow regions and edge information from the edge detection method that was presented in section 4.5.

Due to its simplicity, high speed of operation and ease of implementation, thresholding is a very powerful method for use with digital image processing when the image is segmented and the objects are separated from the background. The method used in this section is based on automatic global multi-threshold values to segment the grey levels extracted from 2DVPI images into the objects and their backgrounds. Generally, the thresholding technique from the histogram of an image is based on the assumption that there will be peak values identified from a histogram of the grey values from the image and the other values will be clustered around them, so if the peak value can be identified then the threshold can be identified [189-190]. This means that the threshold can be identified very simply, if there is an identifiable peak to the histogram, as this is all that needs to be detected. However, in many cases, interesting structures and objects within an image may occupy a small percentage of the background i.e. noises or other occluded objects due to several peaks of very unequal amplitudes separated by a broad contains only one peak. Therefore, the threshold method based on the histogram of one image is limited.

The above difficulty can be overcome using the approach presented here which is based on the selection of the threshold as determined from histogram edge pixels. In the grey level histogram there should be a single peak contained within the pixels related to the neighbourhood of an edge that corresponds to the boundary of the object. Therefore, as long as the edge identification was accurate it is possible to use the grey levels in the boundary to form the threshold that enables the separation of the object from the background. In this process, the edge detection has a very significant influence on the techniques quality, and when generating the grey level threshold the edge position accuracy and edge thresholding process are both extremely important.

Previously, the use of edge information when determining the threshold in a range of algorithms has been suggested [191]. However, the use of these algorithms is limited in automatic systems as the threshold value has to be set manually. The following section details the process of using the results from section 4.5 related to the use of multi-quantized local histogram edge detection in the automatic segmentation of a 3D object in a scene.

6.4.1 Edge Based Multi-Quantized Local Histogram Analysis

In the majority of processing applications, before feature extraction and object segmentation occurs, edge detection will be used as a fundamental tool. This is when objects have their outlines defined to separate them from the background. Most often the edge will be detected in the spatial domain as this generates better results while requiring less computational resources. This work used Multi-Quantized Local Histogram (MQLH) edge maps with multiple thresholds to obtain a single feature-edge map. In the process, a single feature-edge map is generated by integrating discontinuous edge segments from a higher resolution distribution (σ_r) and those obtained from a lower resolution at full-width-at-quarter-maximum (FWQM) ($\sigma_r + d$). The strong edges, which are set by the first starting point (σ_r) (low threshold) and the weak edges, which are set by the second starting point ($\sigma_r + d$) (high threshold) are presented in Figure 6.6.

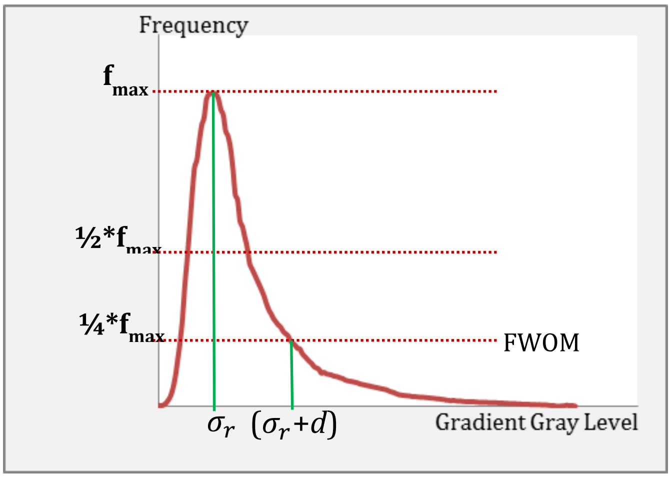


Figure 6-6 Illustration of the gradient magnitude histogram of the VPI showing the two different values of σ_r called the quantizer starting point that identifies the estimated noise levels.

The point at which the quantizer started (σ_r) was set so that it was the same as the value for the grey noise as estimated from a smoothed histogram distribution of the thinned gradient magnitude image. The assumption is that all values less than or equal to the estimated value will be noise while those that remain are the significant features as well as some noise. This process needs to be repeated with increased values of σ_r to remove more noise as all of it is not removed in the first pass.

It was shown from the preliminary experiments that noise reduction could be achieved when the estimated σ_r value was increased by moving it a certain distance (d) to the strong end of the gradient magnitude histogram. The value for (d) is generated automatically from the σ_r value that matches the grey level at the FWQM of the distribution. From this, two quantized images are produced and no important edge segments are lost. From these, many binary images with different edge resolutions are generated from the non-overlapping segment block quantized images when using a single scale local histogram analysis. If the local histogram presents a single peak it will be labeled as a background block and all pixels within it are set to zero; while, the histogram will have two peaks, be bimodal, if there is an edge segment in the block. From this a final feature-edge map is obtained using the edge combining process. The boundaries of objects are located from the edge pixels and then developed into a region that can be used to separate the objects and background in the extracted VPI.

The MQLH edge maps method is adaptive as based on the image's noise level, which is an attractive feature. This is due to the required parameters being generated from each image, so that the noise reduction process is made adaptive depending on the noise present in the given image without the need to increase the size of the local histogram smoothing operator. The method also reduces the computational resources needed as a minimum size 1D Gaussian operator with size W = 3 is used. In addition no parameters need to be supplied as all the required parameters come from standard deviations of the Gaussian smoothing filter and all the parameters are fixed via the edge extraction process.

6.4.2 Labelling Contour Process

When the edge map has been detected all the feature-edge pixels are marked as "1" and the background pixels become "0". The edge pixels are most often located at transitions between two different regions and so can be used a seeds from which the growth of the region around them can be initiated. This means that the gray level intensity in a relatively homogeneous object drops in between the shapes of those objects. Generally the objects will be lighter than the background, and between the dark and light areas there will be a transition region. This transition region will have intermediate grey levels and these levels correspond to the edges of the objects. This grey boundary will enclose an object where the pixels will have a greater value than those of the boundary.

For each object in the image there will be a contour that is a continuous line that separates the object from the background with a grey level value that is between the values of the object and background. This means that the 3D interactive-map generation approach can use the contour edge pixels as seed points from which the boundary can be grown by adding more pixels with similar values relative to the background. As this method only uses FED to initialize and grow the region rather than using a homogeneity criterion it is advantageous compared to other methods. These are the same feature-edge pixels used in Chapter 4 (section 4.3) to estimate the depth map, and simulation results showed that they have the potential and consistency to generate a 3D object map from the H3DI technique. The automatic seed point extraction process is presented in Figure 6.7.

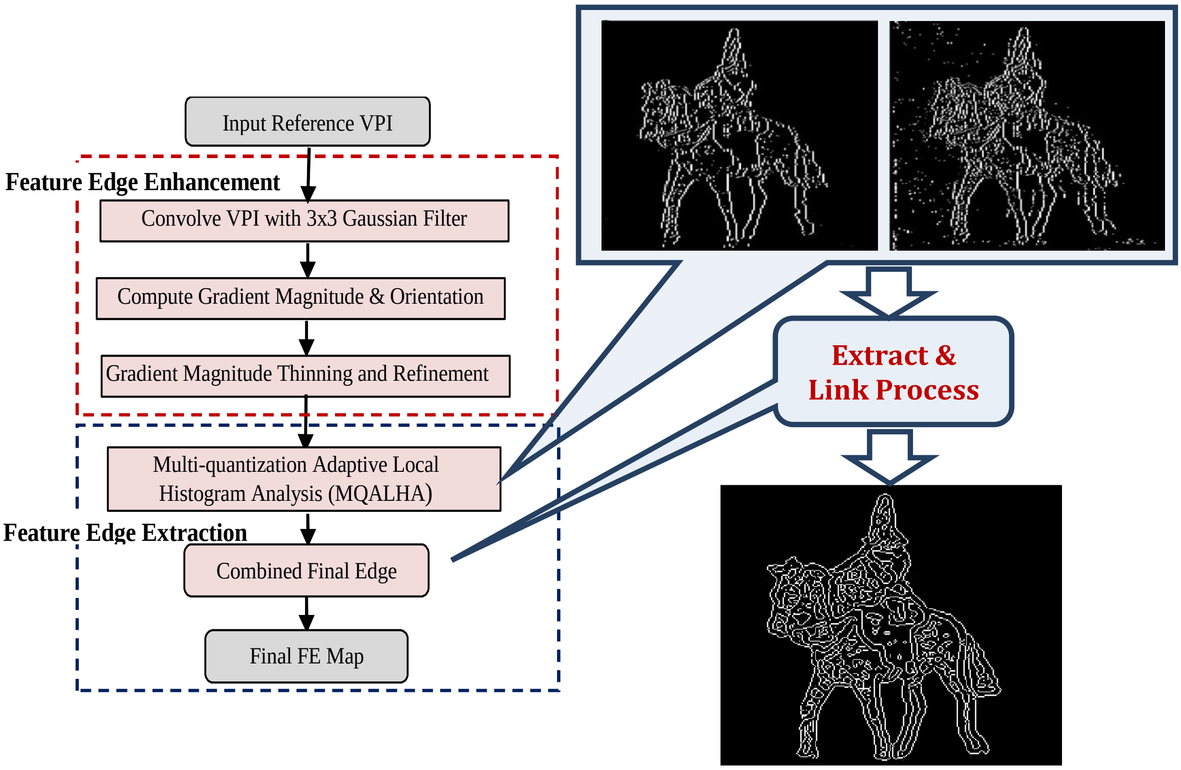

Input Reference VPI
Feature Edge Enhancement
Convolve VPI with 3x3 Gaussian Filter
Compute Gradient Magnitude & Orientation
Gradient Magnitude Thinning and Refinement
Multi-quantization Adaptive Local Histogram Analysis (MQALHA)
Feature Edge Extraction
Combined Final Edge
Final FE Map
Extract & Link Process

Figure 6-7 Illustration of the automatic seed points extracting process.

6.4.3 Region Growing Process

Two popular techniques that are commonly used for image segmentation are region growing and feature-edge detection. When region growing is used with the segmented images, similar pixels are grouped into regions starting from seeds, while gradient discontinuities are used in edge techniques to segregate the pixels. In region growing the seed pixel has compatible neighbouring pixels (those with similar properties) added so that a region is formed based on predefined criteria [192]. Region growing is the preferred technique as it has benefits over edge detection such as being more robust against low contrast problems and providing a solution to connectivity issues [193]. Recently there has been much encouraging work on automating the seed point and threshold value selection. Both region growing and feature-edge based segmentation methods, have positive and negative aspects, and it has been suggested that combining both methods would generate much better results over all [194]. It has also been suggested that in region growing, already-extracted edges and completion of these edges can be used as well as pixel and neighbourhood similarity in region joining decisions [195].

In this work region growing is based on the extracted feature-edge information that is used to select pixels automatically and guide the growth. The region growing starts from a seed point and selecting the correct point is very important for the segmentation result. The seed point has to be selected from within a region of interest (ROI), otherwise the final segmentation will be definitely incorrect. Recent studies have focused on integrating region growing and edge detection; for example this hybrid method using mathematical morphology is presented in [196]. A fuzzy technique was used to automatically determine the threshold that controlled the region growing process. When used with the 3D segmentation of MRI brain images the 3D segmentation method gave results that were superior to those obtained from a single technique used in the segmentation. A combined region edge preserving smoothing algorithm has been developed that removes the errors commonly encountered when using the methods individually [197]. A criterion has been suggested that would determine if the growing process should continue or terminate when each new edge pixel is encountered [198].

In this proposed technique, the following equation gives the syntax for the function to perform basic region growing:

$$VPI_{output} = Regiongrow\ (VPI, \alpha, \beta) \qquad (6.5)$$

where, the segmented output image is VPI_{output}, an integer value is used to identify the parts of each section, the image to be segmented is VPI, the input feature edge map has 1's where all the seed points are and 0's in the other places is α and the array is β, in which there is a value for each location within the VPI. A noteworthy form of boundary-oriented pixel labelling for an automatic seed growth algorithm has been established in which the object is determined via dilating its boundary step by step [199]. Via the use of an 8-connection aggregation window, it is possible for this edge-oriented seed pixel determination method to maintain parallel region growth.

When the seeded area is first being grown using the seed region growing (SRG) procedure, the cues related to the boundary pixels and the centroid for a certain area are equal. When this method is implemented, all the regions commence growth from their seeds simultaneously, which is a major advantage. The assumption that underlies the algorithm is that when an unlabeled pixel is encountered that is next to a 3×3 sized aggregation window it will be similar to the adjacent boundary feature-edge pixel (BFEP) and thus it will replace the BFEP region once merged with it. Then the pixel's intensity should be compared with that of the current test pixel and it should be merged into the region that it is most similar to, thereby becoming the new boundary pixel. To grow the regions, all unallocated neighbouring pixels should be compared with the regions. It should be noted that the results obtained will be based on the specific choice of seed pixels.

6.4.4 Proposed Methodology and Implementation

This work aims to identify, extract a depth information map and produce a segmentation of a 3D object in the scene from a H3DI system via the use of 3DIM. The 3D object segmentation method is based on the methods of 3D information map technique and that of 2DVPI segmentation to form a combined segmentation technique. Within this, the 2D segmentation used in this approach is a hybrid method. This is a combination of edge based and region growth methods to produce the 2D foreground map; and, this hybrid method also contains a filter that enhances object separation by clearly highlighting the boarder. Figure 6.1 presents the overall framework of this method where the depth map is merged with the VPI segmentation module's output to generate the final form. Using the previous method as a base, this work explains a simple segmentation method to highlight the objects and separate them from the background. The edge map, produced from the FED algorithm, is used to determine the depth of the objects within the VPI displacement with an equation, and this produces the mathematical relationship that is present in the interaction of the object depth and correspondence VPI displacement. As explained in Chapter 4 section 4.3, the adaptation of the multi-baseline technique is implemented and it uses the fact that when there are many VPIs of the same scene there is extra information. The same extracted features that identify the object boundaries are also used as seed points for the region growing.

6.4.5 Experimental Results

The 3D object segmentation method was implemented on real-world H3DIs with the feature-edge detection stage being undertaken using the edge detection algorithm from Chapter 4 that produced reliable features referred to as "feature-edges". The object EFs maps produced are distinctive and robust when there are changes in the VPIs, as they have strong local feature information. The EFs can be used to both highlight geometric constraints and provide strong feature correspondences between the VPIs that form the depth estimation, while also being the initial seed locations. A range of real data H3DIs were used to test the algorithm. The "tank" image has a poor texture while the "horseman" is noisy with various levels of texture. Figure 6.8 presents the outline generated from the EFs extraction, which is the first stage. Binary object outline EFs maps were produced from both images. The white represents the object feature-edges, as denoted by "1's", and the black is the background, as shown by "0's". The object boarders are very clear and used to enhance the separation of the object. As can be seen from the results, a suitable object contour is generated from which the fine details can be identified. For each H3DI, the center extracted VPI is presented in Figures 6.9 (a, c) while Figures 6.9 (b, d) shows the 2D foreground masks that are produced in the second stage that are the growth regions.

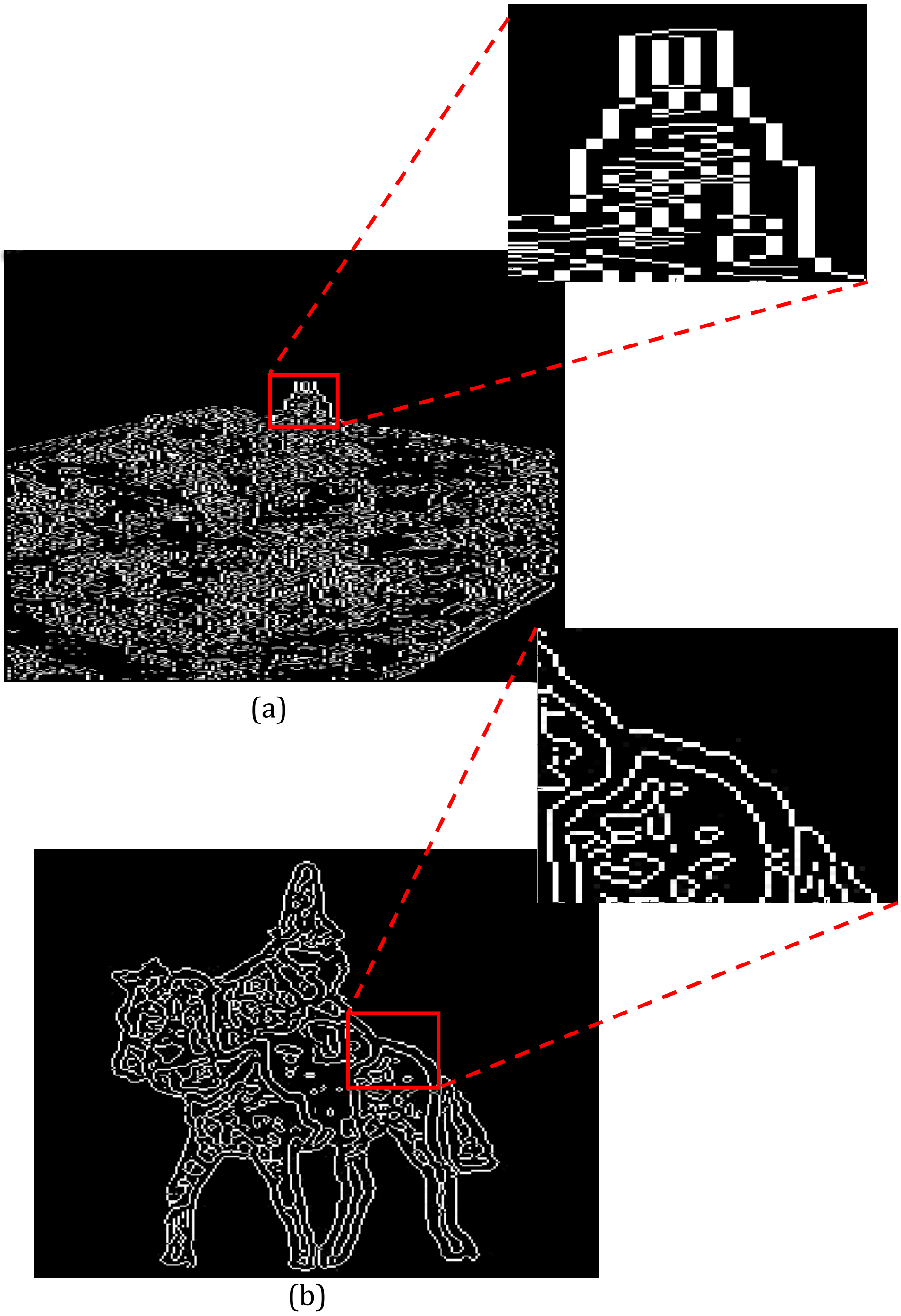


Figure 6-8 Illustration of the object outline seed points map, (a) Tank and (b) Horseman, and their magnified section.

As all the contours of all the objects can be observed after the edge map cleaning process that takes place in the first stage, the segmentation process has been successful and it can be used to enhance the object separation. Now the binary object segmented map can be used as a filter to ensure the segmentation and separation of the 3D objects from the scene, which is also referred to as 3DIM.

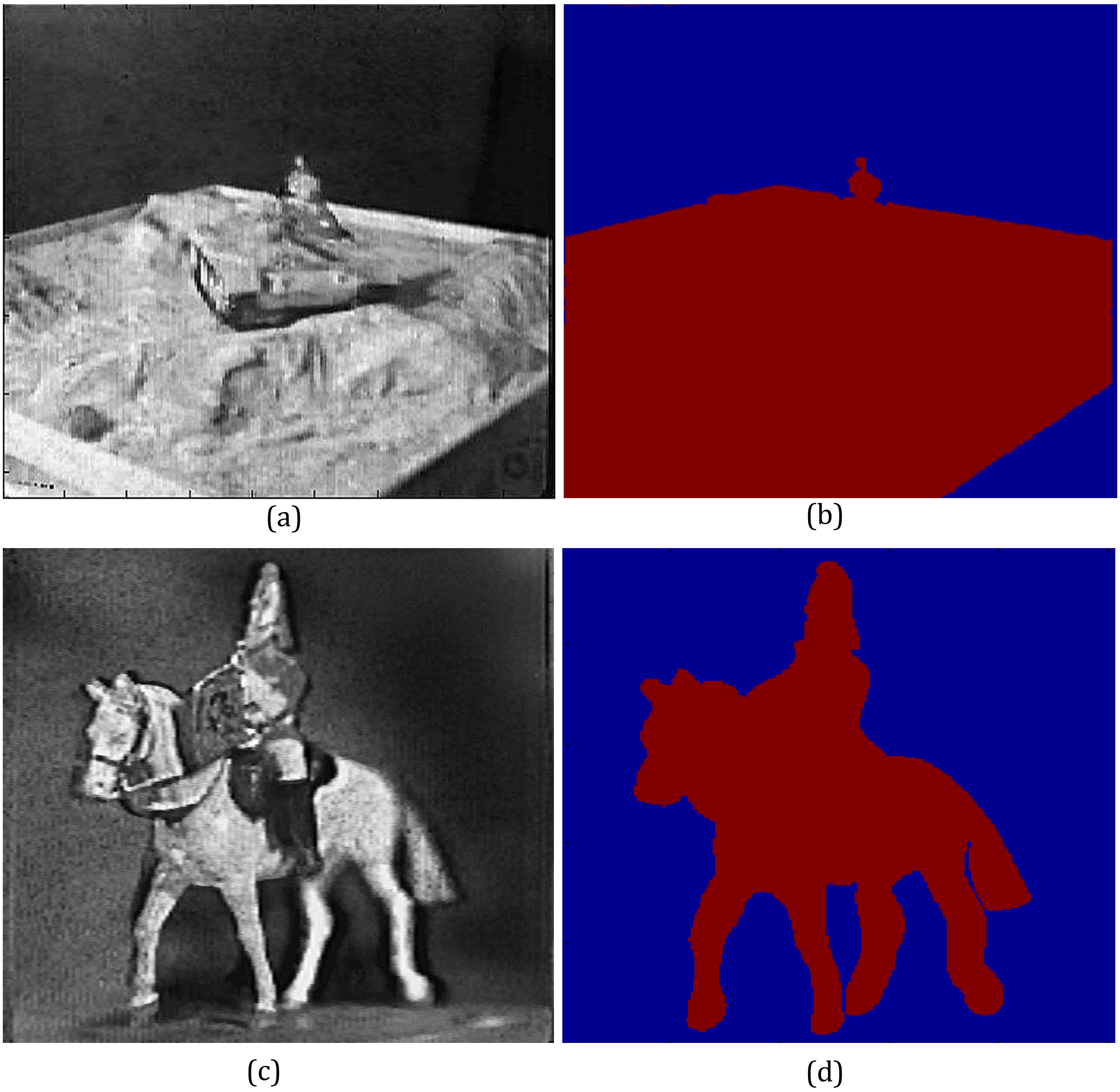


Figure 6-9 Outline of 2D segmentation algorithm, representing the foreground mask results, (a) Tank and (b) Horseman.

Further tests were performed with the algorithm on two type images; real and C.G. H3DIs, that were recorded using an alternative cylindrical lens array to the one used previously. It can be observed from the results that the algorithm was able to determine a surface with smooth contours and identify objects even in images with poor texture qualities. Binary masks from different VPIs with various textures are shown in Figure 6.10, and these can then be used by the algorithms to undertake a useful task.

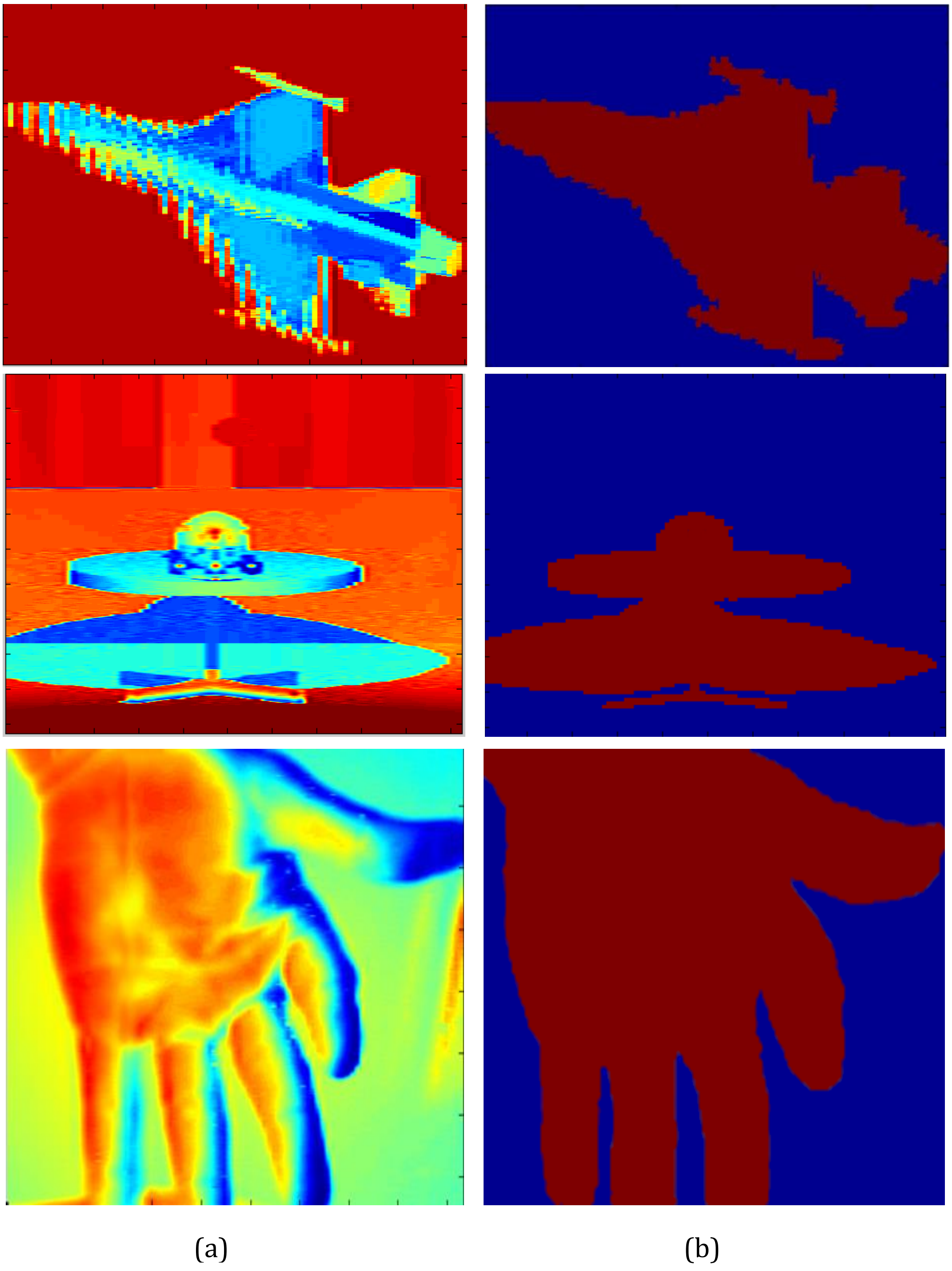


Figure 6-10 Results of the proposed 2D segmentation algorithm. (a) The first column represents the central VPIs, the first and second images are the synthetic VPIs "plane and table", respectively, while the third VPI is real data of "palm", (b) foreground mask results obtained from the extracted VPIs. It is worth pointing out that the VPIs are displayed in colourmap "jet" ranges from blue to red due to uncontrolled changes in luminance that prevent meaningful conversion to grayscale for display and different regions of the image are mainly distinguished by this colourmap [200].

The part of this work that is novel is when the generated 3DIM automatically integrates the depth estimation technique and the FEs segmentation technique. The depth map estimation process is related to the use of the final FEs map to generate the depth through extraction VPIs and disparity (as in section 4.3) when producing a 3DIM. When the depth map and the VPI segmentation module's outputs are joined, this will be the final output. It should be noted that the results presented in Chapter 4 Figure 4.18 and Chapter 5 Figure 5.14 are the preliminary results from the 3D object segmentation algorithm as shown by the 3DIMs. All the 3D objects generated by the H3DI system have been generated as required by the 3D segmentation process.

When the final 3DIMs are studied it can be seen that the majority of the object boundaries have been correctly identified, and that there is minimal noise. The main reason for this is that the auto feature-edge detection algorithm has performed well and provided efficiency and robustness for the varied domains. In the first it set and selected the FEs for the 3D depth map estimation. Then in the second it enabled the separation of the objects from the background in the VPIs.

6.5 Evaluation of the 3DIM

When evaluating whether the segmentation algorithm meets its targets, it is vital to be able to determine the quality of the output result. The output's targets are related to specific applications. Some situations may need very fine details to be identified while in others such a fine level of detail may not be required. The 3DIM presented in this work is used to find the places where there are objects with a suitable level of precision that they can be separated from the background. To evaluate whether a given segmentation algorithm reaches the application goals, both objective and subjective evaluations have been used to determine if the 3DIM generation algorithm meets the requirements. This section aims to present and analyse the output results of the various tests performed according to the foreground object extraction techniques from SRVPI, which was presented in the previous chapter. The objective and the subjective test scores are used individually to evaluate the performance of the two foreground object extraction techniques, the 3DIM creation and SRVPI generated process using the numerical data collected as well as through visual comparisons algorithms.

6.5.1 Objective Evaluation: Scores and Analysis

The proposed foreground object segmentation techniques are evaluated quantitatively to compare the performance of both techniques using synthetic and real data VPIs. Then the outputs from the methods presented are compared to that of state-of-the-art segmentation algorithms. Previously, foreground objects were compared to "ground truths" with regards to false negative error (FNE) and false positive error (FPE), for the foreground pixels missed and background pixels misclassified, respectively [201]. When considering the results from various environments the foreground objects were determined to not be accurate enough. An Intersection-Over-Union (IOU) similarity measure was developed to combat the above problem where the "detected region" and "ground truth" were used to test the foreground segmentation [202]. The following is the definition of the similarity measure:

$$IOU\ (DR,\ GT) = \frac{DR \cap GT}{DR \cup GT} \tag{6.6}$$

where, the detected region and the related ground truth are represented by DR and GT respectively, while the false positive and negative errors become a single parameter.

This method reviews the function of the two foreground object segmentation techniques used for forming the 3DIM, with the IOU similarity measure that is presented in Eq. 6.6 used. IOU similarity (or Jaccard similarity) has been the standard measure since 2008 to determine the level of pixels that are incorrectly labeled [203]. Manual segmentation was used to separate the "ground truth" foreground objects pixels so they could have their values determined, and various examples of the ground truth binary masks and segmented images are shown in Figure 6.11.

For the IOU the values will be in the range of "1" to "0" where a value of 1, shows that the DR and GT are exactly the same, and then the range of values to 0 show various levels of dissimilarly. Using the two methods, Eq. 6.6 has been used with the results obtained, and the results were compared to those obtained from existing segmentation techniques. The IOU similarity measure was used with the active contour based segmentation model, as this has a clear curve for the object [204].

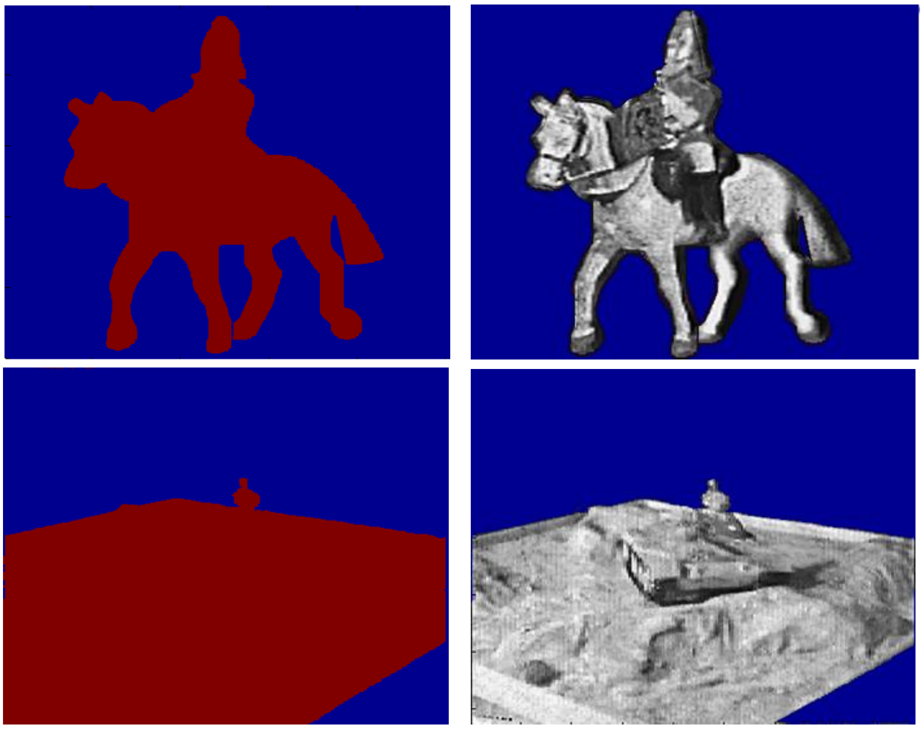
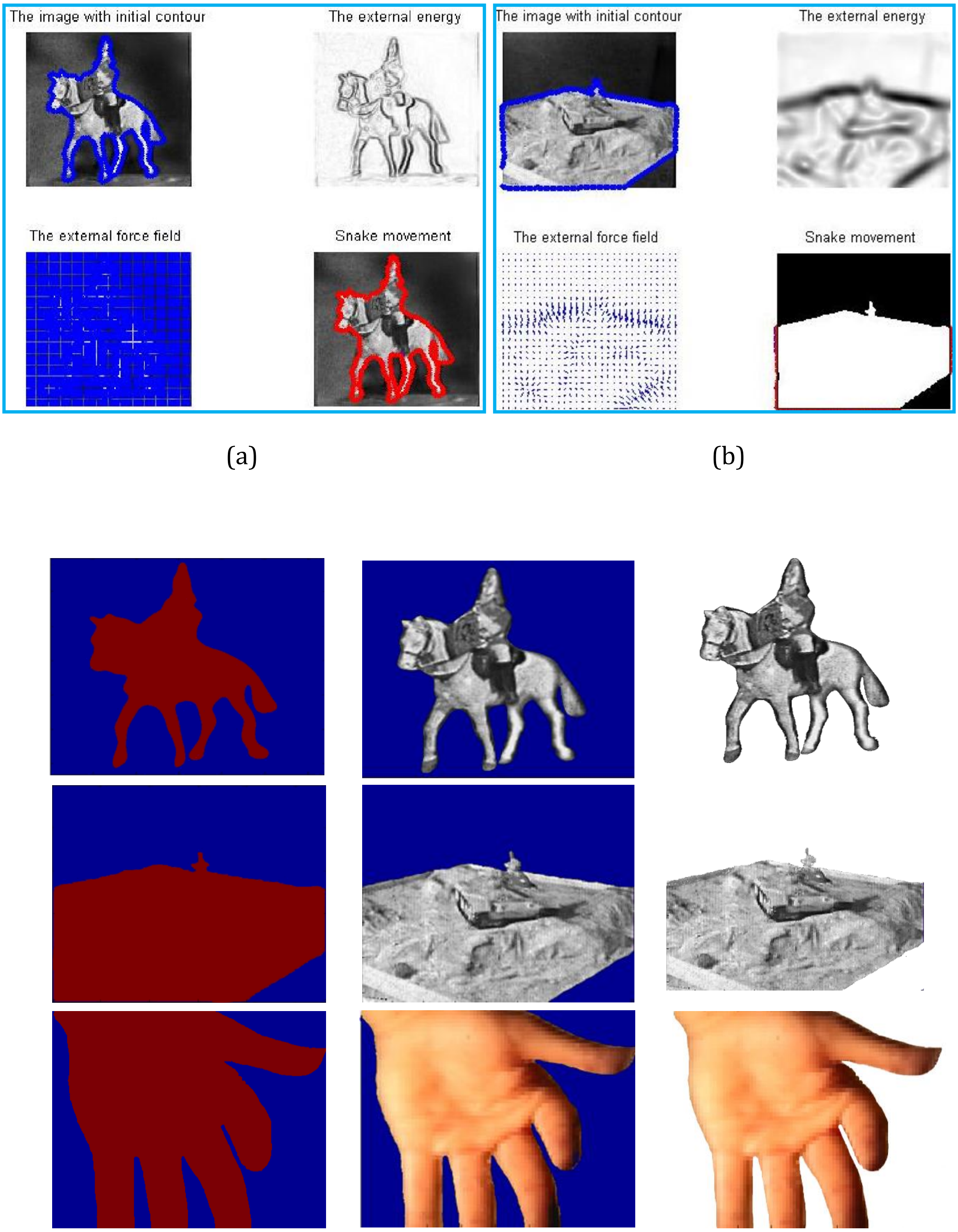


Figure 6-11 Example images considered in the test, the left column shows the binary mask of “horseman and tank” segmented manually from the original images, the extracted ground truth object image using the binary mask is shown in the right column.

An active contour based object segmentation algorithm [205] has been applied as it has previously found uses in a wide range of areas, such as medical imaging, motion analysis and image registration. When using the method it is necessary to manually set the anticipated contours.

Examples of the active contour process [205] using 250 manually identified points that are located close to the object edges are shown in Figure 6.12(a, b), while the binary masks and segmentation objects that are formed are shown in Figure 6.12(c-e). The values obtained from similarity measures for the contour algorithms from real and synthetic images are presented in Table 6.1. From the tested images, six that were VPI real or C.G. were used further in this section. The images contain various textures and changes in the background to enable the methods to be evaluated, and considering the challenges, both methods provide good results. Table 6.1 shows a comparison of the methods similarity measures, which are illustrated in Figures 6.9(c, d), 6.10(b) and 6.12.



(a) (b)

(c) (d) (e)

Figure 6-12 Process and output of active contour model on some test VPIs using [205] model. (a) “Horseman” active contour process used to locate object contour to extract object in the scene, (b) “Tank” active contour process shows the generation of binary mask of “Tank”, (c) active contour method binary masks results, (d, e) segmented object in the scene and their extracted objects.

The new region-growing based feature-edges method shows an impressive result compared to the intensity distribution method. This was due to the information related to the surface structure that came from the feature-edge maps via the use of the MQLH edge maps at a range of thresholds. In the intensity distribution segmentation method the object mask and its quality is related to the morphological structuring elements and their size. This was used to probe the object's shape in the image. This parameter needs to be determined manually.

When comparing with the manual object segmentation (active contour object segmentation), the second proposed method had better performance related to four of the six images, and was only limited in the other images (kidney and hand) due to the fact that real-world textures are often not uniform. The results show that as far as the shape and boundaries were concerned the recognition was correct (Figure 6.13). Figure 6.14 show the "horseman" image and the results from the use of the two proposed automatic segmentation techniques and the manual active contour model as well as the ground truth results.

Table 6-1: Quantitative evaluation and comparison of IOU values.

VPIs	Real Horseman	Real Tank	Real Hand	Real Kidney	C.G. Table	C.G. Plane	Average
Active Contour Model	0.907	0.789	0.813	0.815	0.711	0.827	0.810
1st Proposed Algorithm [section 6.3]	0.612	0.743	0.435	0.578	0.533	0.480	0.563
2nd Proposed Algorithm [section 6.4]	0.918	0.958	0.661	0.709	0.782	0.912	0.823

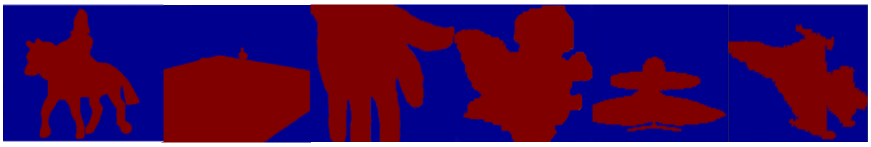


Figure 6-13 Examples of IOU values of each individual tested image obtained with different methods and the result obtained from region-growing based on edge detection i.e. second proposal.

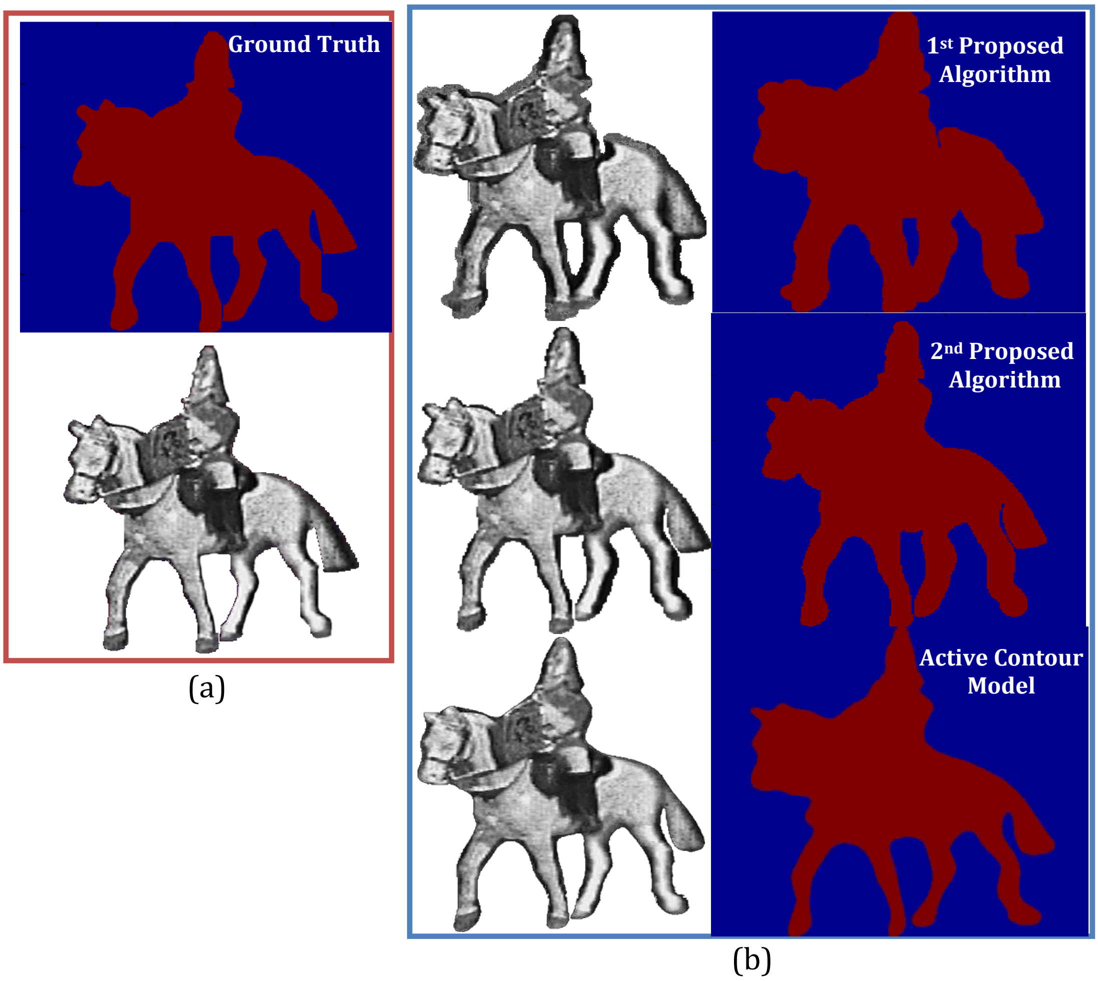


Figure 6-14 Resultant foreground masks of “Horseman” obtained from, (a) manually, (b) first and second rows results from the two proposed automatic segmentation techniques and the active contour model results display in the third row.

6.5.2 Subjective Evaluation

The qualities of proposed 3D image segmentation algorithms are evaluated in this section using subjective criteria as estimated by viewers. A different evaluation method must be used, where the human factor must be taken into consideration. Therefore, the experiment must be conducted with the help of human participants in a subjective manner. The ideal method to achieve this is an explicit subjective experiment where the results are directly rated by the viewers. Determination of the visual perception of the overall quality via subjective evaluation of the 3D segmentation is required. For the 3D contents to be successful when used by the public costumer market, visual quality must be compared to conventional standards to guarantee a strain free viewing experience [206].

The aim of the assessment is to determine the participant's ratings of the 3DIM including the areas of "SRVPIs quality" and "depth quality". During the 3DIM's evaluation the depth perception was measured via determining the relative positions of objects and the depth layout quality that was determined by the depth information being realistic or plausible. The impact of depth information on the perceived 3D image quality is one of the main issues that has to be investigated. Recent studies show that depth is not related to the perceived 3D effect. However, other studies point out the importance of depth for quality perception [186]. A previously identified subjective analysis methodology was utilized in [207, 208] to evaluate the 3DIMs, where there is not a reference benchmark "ground truth depth map" available for the H3DI system. The method was based on the Single Stimulus Continuous Quality Scale (SSCQS) that can be used when there is no reference to refer to [208]. In this approach, human subjects evaluated a range of depth maps that had been segmented by four different algorithms. A five point scale ("5-Excellent", "4-Good", "3-Fair", "2-Poor", "1-Bad") is used to state the quality of experience (QoE) of the depth maps. While, the double stimulus continuous quality scale (DSCQS) method was used to evaluate the SRVPs that had been formed from multi-low resolution VPIs. The reference VPIs of the test images are used, therefore the observer can determine the relative quality of the distorted images. The scale of ("5-very high", "4-high", "3-medium", "2-low" or "1-very low") was used for the SRVPI quality [209]. The pooling of the votes into a mean opinion score (MOS), which provides a measure of subjective quality on the media in the given test set was used. Assessment sessions were only half an hour in length so that the viewers would not become fatigued. The viewer was alone when viewing the images and there were no external noises that could distract the viewers.

6.5.2.1 The Research Methodology

During the subjective quality assessment the following steps were taken:

1) **The test session:** Nine H3DIs (6 real-world; three OH3DIs and the other UH3DIs as well as three C.G. data: one OH3DI and two UH3DIs) were used. Two scales (2DVPI and depth rendering quality) were used in each session to rate each method (two scales by four methods) so that six images were rated in total.

2) **Equipment**: The assessment was conducted in a room that matched the recommendations for subjective assessment of visual data as set forth by ITU-R BT.500 [209]. For the assessment two 32" display monitors with native resolutions of 1920 × 1080 were used.

3) **Viewers:** The test was conducted with 32 participants (14 female and 18 male), who were selected randomly from university international students and staff members. The average age was 31 years old and overall range: 21 to 41. Also, the overall experiences of the participants with 3D technology and image processing have been considered. Therefore, based on the participant's experiences they are divided into two groups in order to assess perceived overall image quality and perceived depth. The first group of the subjects Group 1 (G1), have expert knowledge of both 3D technology and image processing. The second group, Group 2 (G2) do not have expert knowledge of neither 3D technology nor image processing. The purpose of this division was to obtain a more narrow and reliable analysis for the assessment regarding the Mean Opinion Scores (MOS). The availability of MOSs is fundamental to enable validation and comparative benchmarking of objective quality assessment systems to support reproducible research results. Table 6.2 illustrates the participants demographics where 55.25% of the participants were male and 43.75% were female.The participants had normal vision, or corrected so that it was equivalent to normal. The viewers started the main experiment individually and were requested to be as accurate as possible in their judgment of each viewing session. They were asked to complete the questionnaire based on their judgment.

4) **Stimuli:** Six UH3DIs were utilized in this assessment, which were as follows; four real-world namely: horseman, tank, hand and kidney, and two synthetic images: table and plane, were used. In addition, for the generalization capability aspect, the evaluation involved two real-world and one C.G. OH3DIs.

5) **Procedure Instructions:** Written instructions (Appendix C) have been provided to the test subjects that explained what was required. To ensure understanding and compliance the experimenter also reiterated the instructions with the subjects. The subjective test methodology was introduced and the depth map levels were presented using training stimuli during a training session. The viewers were told that once scores were recorded they could not be altered. On the session of depth map

quality, they considered four 3D depth maps from different algorithms while on the session of image resolution quality they were scoring on four VPIs produced from different methods; the sampled pixels, proposed algorithm, Vandewalle et al. and Keran et al.'s algorithms (see section 5.3) were used for each test image. The participant was presented with all nine test images so they could give their opinion according to the ITU-R BT.500 five point quality scales [209]. Four produced images and four estimated 3D depth maps from different algorithms were displayed on the left and right screen, respectively for each test image for 3 minutes. A continuous scale was used to rate them: [bad]-[poor]-[fair]-[good]-[excellent] for 3DIMs and [very low]-[low]-[medium]-[high]-[very high] for VPIs resolution. A specifically designed questionnaire for this test is given in Appendix D.

A numerical scale (0-100) was also used in the evaluation to present exact values. Using this scale a stimulus that gave high sensations would be 100 (excellent or very high) with a bad sensation being scored as 0, and in-between values scored with a comparable numerical value. However, when using any scale it will be subjective and vary between participants; therefore, the steps in Figure 6.15 were applied [210]. These transformed the raw subjective scores in to a Mean Opinion Score (MOS). Then for each individual test k, the perceived quality was determined:

$$MOS_k = \frac{\sum_{l=1}^{N} S_{lk}}{N} \tag{6.7}$$

where, for each test k the mean opinion score is MOS_k, the amount of viewers in N and S_{lk} is the score ($\in [0, 100]$) of the viewer l of the k. Then it is possible to calculate the average of the MOS_k for each category and each algorithm of the two assessments types (3DIM and VPIs quality). Each of the two assessments results were performed separately according to the steps described in Figure 6.15.

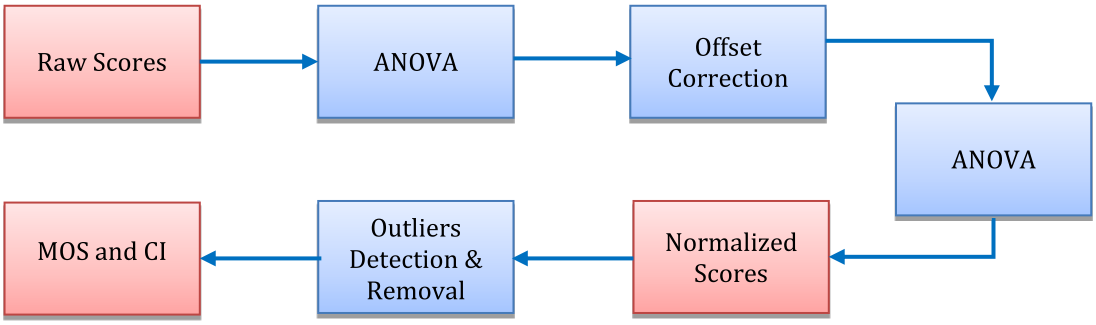


Figure 6-15 Flow chart of the processing steps applied to the subjective data in order to obtain the MOS values [210].

The process set out in [210] for processing the raw subjective scores is as follows:

1. *First phase:* Via the use of ANalysis Of VAriance (ANOVA) it was determined whether or not the score data would need to be normalised. This showed that there were large differences between the participants' scores, which showed that they applied the rating scale in different ways.
2. *Second phase:* All the scores were normalised via an offset mean correction so that there was a viewer-to-viewer correction.
3. *Third phase:* With regard to the normalised scores, suspected outlier viewers were identified using the guidelines described in Section 2.3.1 of Annex 2 from [209].

Table 6-2: Analysing the collected data of participants.

Demographic Variable		Participants	Percentage	Total
Gender	Male Female	18 14	55.25 43.75	32
Age range (yrs.)	21-28 28-41	15 17	46.875 53.125	32
Educational level	PhD MSc BSc	18 8 6	56.25 25 18.75	32
Group 1	Expert	15	46.875	32
Group 2	Non-expert	17	53.125	
Group 1 after Refined	Expert	15	50	30
Group 2 after Refined	Non-expert	15	50	

6.5.2.2 The Scores and Analysis

A total of 32 participants voluntarily took part in the experiment. Form the 32 observers, two outliers were found in the results (see Table 6.2). The outliers were discarded and the *MOS* has been determined for each test condition at the 95% confidence level (CI). To measure the statistical reliability of the predicted data, the CIs of $(100 \times \alpha)$ % were calculated on the *MOS* values using the Student's *t*-distribution, as follows:

$$\mathrm{CI}_j = \frac{t(1-\alpha)}{2} \cdot \frac{\varepsilon_j}{\sqrt{N}} \tag{6.8}$$

where, the significance level is α, the *t*-value associated with the required significance level for a two-tailed test with N-1 degrees of freedom is $\frac{t(1-\alpha)}{2}$, the amount of observations (total minus outliers) is N and the standard deviation of the observations j for N observers is ε_j. The α value of 0.05 is used in this situation is comparable to a degree of significance of 95% confidence intervals, and this confirms that there are no statistical differences between the *MOS* values. The overall VPI quality and depth perception are rated by both groups.

1. Analysis MOS of 3D Depth Map Segmentation

According to the collected data, it is possible to compare the performance of the 3DIMs quality metrics using the averaged mean opinion scores. This analysis consists of the four rating methods (first method proposed in section 6.3, second method proposed in section 6.4, active couture model and *C. Wu* algorithm) for depth map quality assessment. The average score given by each participant for the depth maps produced from all the test H3DIs using the four algorithms at 95% *CI* after the refining procedure, is presented in Figure 6.16. It can be observed that the overall *MOS* values of the proposed second automatic algorithm are higher than the three other algorithms. In addition, the *MOS* of this automatic algorithm is nearly equal to the values of the *MOS* for the wide-user manual segmentation algorithm (active contour model).

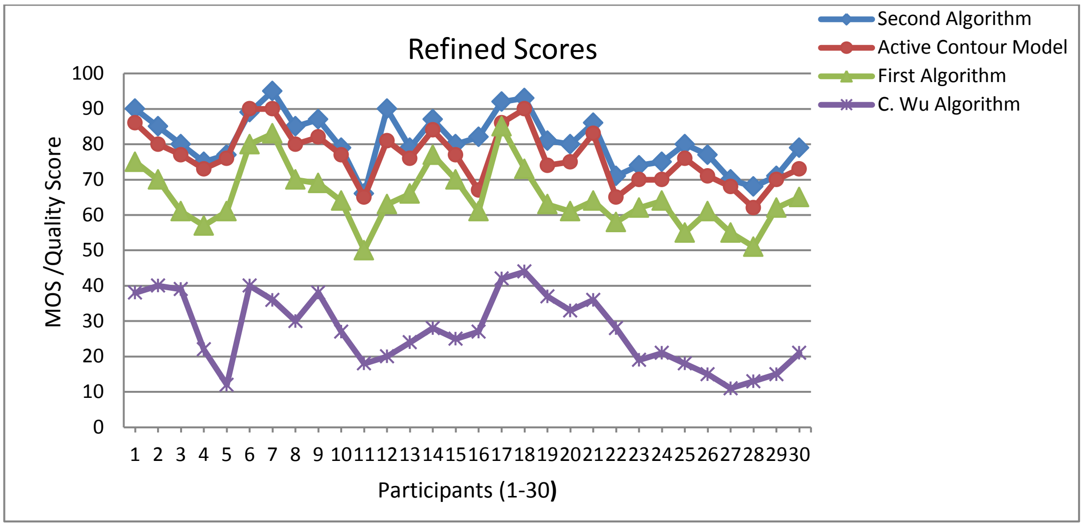
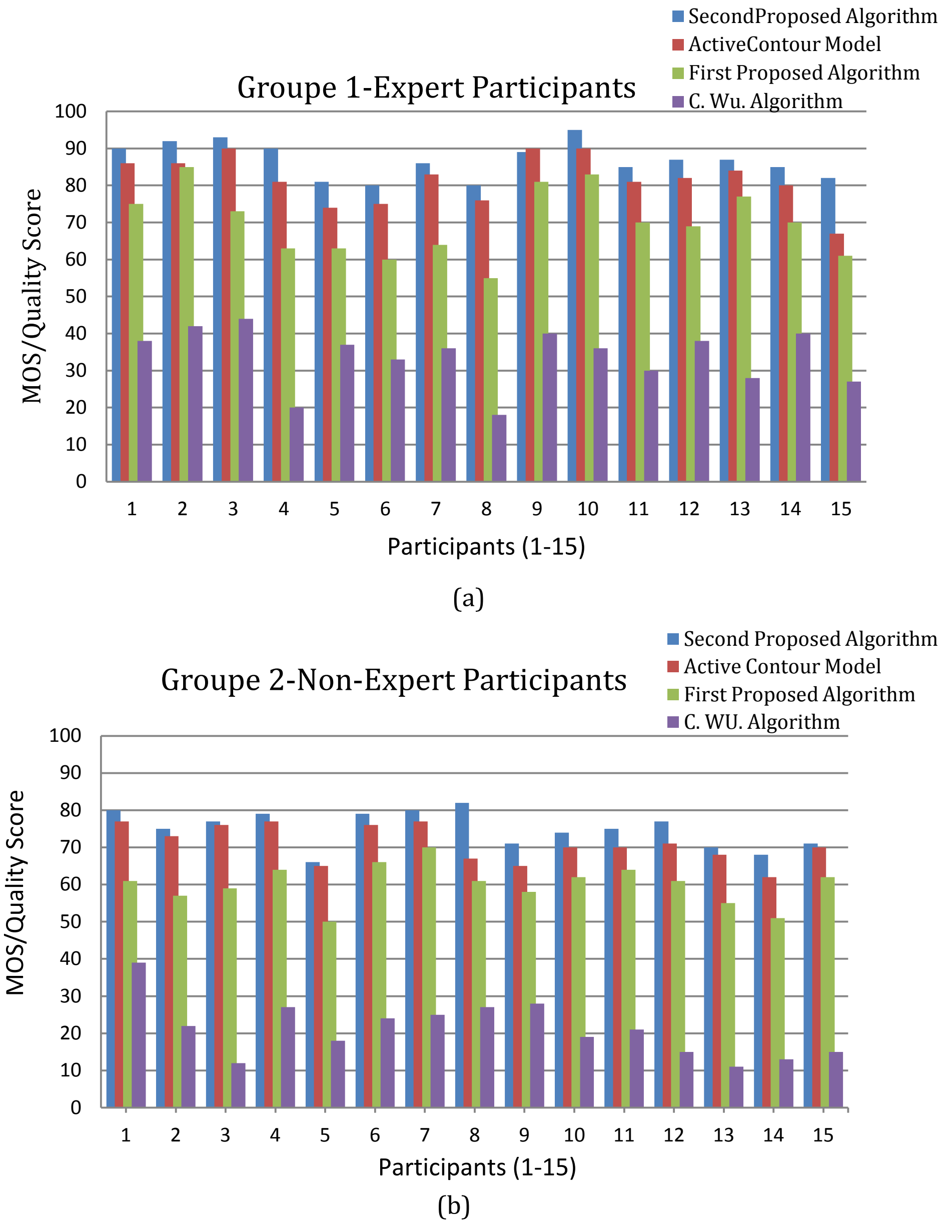


Figure 6-16 Overall average score by each subject.

The *MOS* scores given by the expert participants (Group 1) and non-expert participants (Group 2) are shown in Figure 6.17. This shows that the differences among the *MOS* values of the four algorithms are very significant within each group, and that the evaluations of them for both groups are very similar. Figure 6.18 contains the MOS result for test images produced from the automatic algorithms, 1 and 2, as well as the manual algorithms, active contour model and *C. Wu.* The results are similar for both groups for all test images, but with significant increases in most of the *MOS* rating values from the expert group. Based on the presented subjective scores, Table 6.2 lists the *MOS* percentage of the 3D segmented images, for each method, that are rated average and above (Excellent, Good and Fair) and below average (Poor and Bad). The average *MOS* shows that the *C. Wu* algorithm performs below average. 93.33% of this algorithm's output is rated below average, indicating that the 3D depth map estimated using this method shows, typically, "poor" subjective quality. This is mainly due to the noticeable bulky noises associated with the segmented 3D depth map where the algorithm fails to separate the object from its background. In addition, according to Table 6.3 comparing the totals, the other three methods performed with an above average rating. The second algorithm had the highest percentage (46.6%) of "Excellent" ratings compared with the other two methods, while the first algorithm had the lowest percentage rate. This significantly indicates that the second proposed algorithm outperforms the other algorithms.

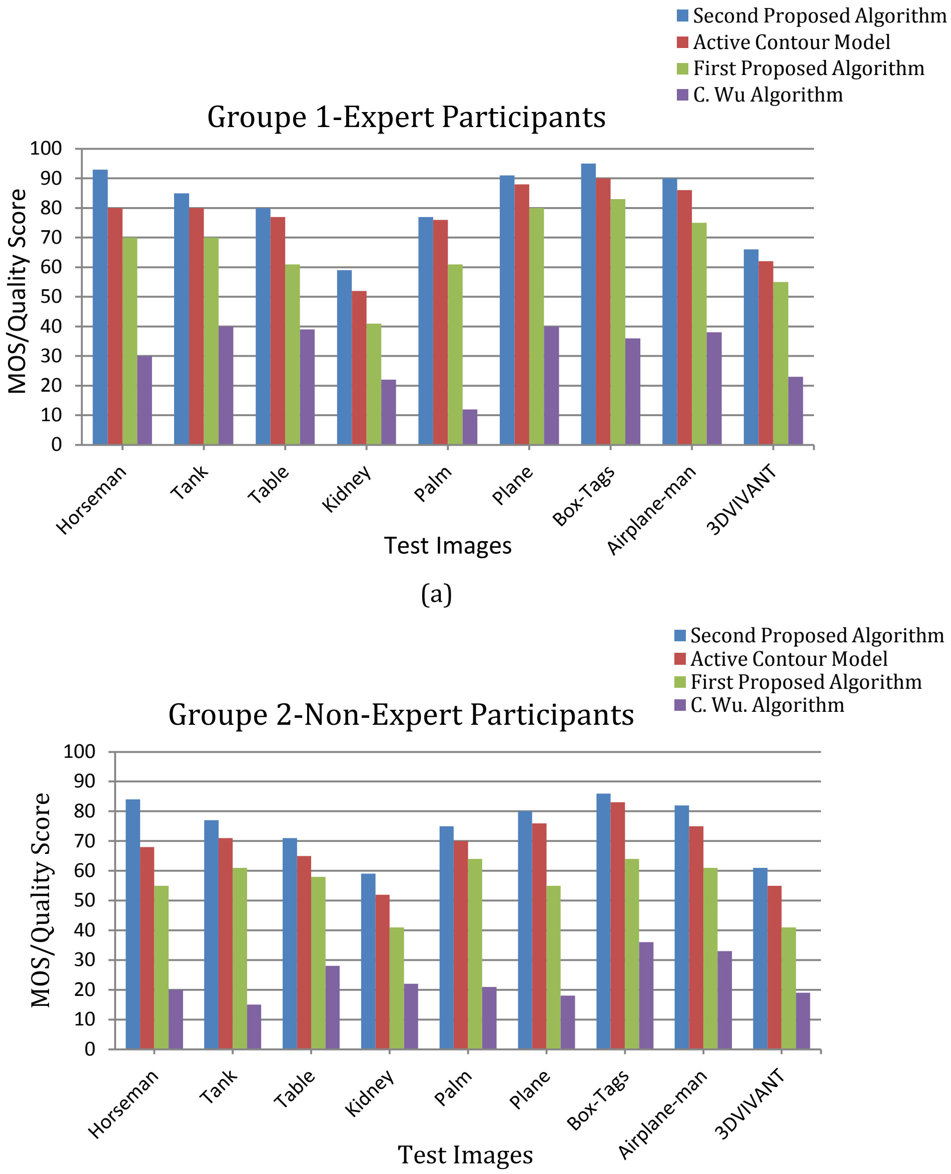


(a)

(b)

Figure 6-17 Comparing MOS scores for 3D depth map segmentation from different algorithms, (a) Expert participants /Group 1 and (b) Non-expert/ Group 2.

As seen in Figure 6.17, the proposed second algorithm outperforms all the others, including the active contour model, which is a manual ground truth method generated subjectively. The objects in the second algorithm are finely segmented, whereas the objects in the active contour model are segmented less. The proposed first algorithm is comparable to the manual ground truth result, and it outperformed the state of the art *C. Wu* algorithm, as shown in Figure 6.17. The evaluation was performed on nine images and the results are presented in Figure 6.17 and Figure 6.18.



(a)

(b)

Figure 6-18 Quality rating results comparing the segmented 3D depth maps produced from different algorithms for nine teste images. (a) Expert participants/Group 1 and (b) Non-expert participants / Group 2.

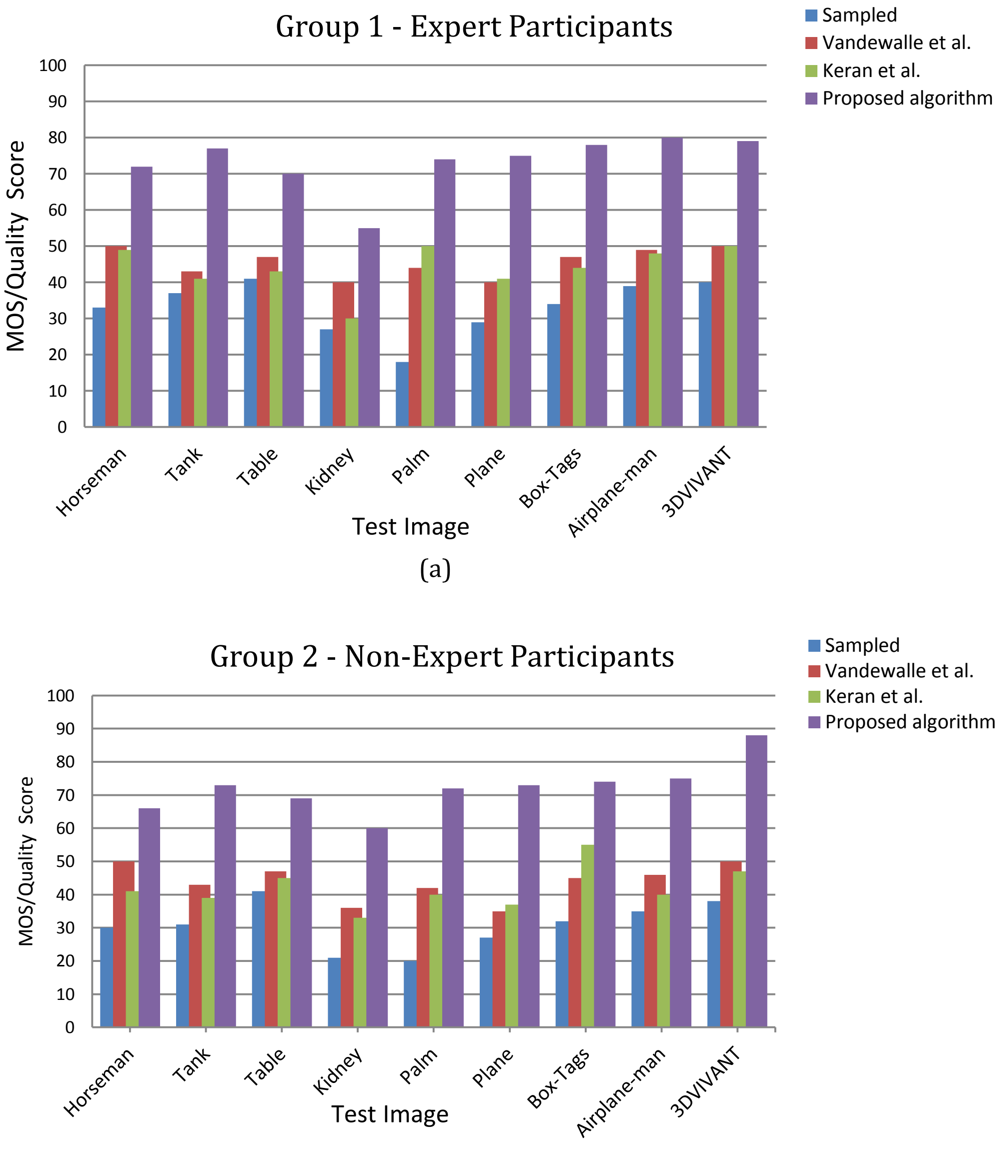
Table 6-3: Percentage of *MOS* ratings attributed for each 3D depth map produces method.

Methods	**Below Average**			**Average & Above**			
	Bad	**Poor**	**Total**	**Fair**	**Good**	**Excellent**	**Total**
1st Proposed Algorithm	0%	0%	0%	26.6%	63.3%	10%	100%
2nd Proposed Algorithm	0%	0%	0%	0%	53.3%	46.6%	100%
Active Contour Model	0%	0%	0%	0%	63.3%	36.6%	100%
***C. Wu* Algorithm**	30%	63.3%	93.3%	6.6%	0%	0%	6.6%

2. Analysis MOS of SRVPI Generation

When using SR reconstruction to form an image, one of the problems encountered is to find a method of determining the image's quality. Image processing experts often use Peak Signal-to-Noise Ratio (PSNR) and Mean Structural SIMilarity (MSSIM) as possible methods, as they can be used in prepared experiments for SR with a synthetic data set. However, in real-world applications, there will not be a known reference image. When using *MSE* error based measurements it is not possible to determine the level of visual improvement in the image. Therefore, a previous study did not attempt to determine the quality of the reconstruction, and just used comparisons of printed images undertaken with the human eye [211].

In the work presented here, to evaluate the performance of the proposed method for generating SRVPIs from a set of low resolution VPIs, a subjective quality assessment is used. The distribution of the two groups under study for the SRVPIs obtained from the proposed and established methods are compared for a group of test images as shown in Figure 6.19 and Figure 6.20. As the results show, it was clear that when comparing Group 1 to Group 2, they both had very similar trends. This indicated that the participants, irrespective of their background in either 3D technology or image processing, are able to judge whether the quality of the image is good or bad in the same manner. However, the *MOS* rating values of the expert group are higher than the non-expert group, possibly due to their work experience in the field of image processing and 3D technology. For the purpose of understanding which methods can produce 2D SRVPIs with acceptable quality or poor quality. The *MOS* average score was generated based on this assumption by grouping together the four relevant methods of generating the VPIs. Table 6.4 listed the *MOS* percentage scale values for the SRVPIs generated from the four methods. According to Table 6.4, there are three methods that were evaluated to have average or above average scores. 83% of the proposed method's outputs are rated as average or above with the "high" scale, while *Vandewalle et al.* and *Keran et al.'s* methods were rated as 77.8% and 55.6%, respectively, at the "medium" scale. 87.9% of the sampled VPI method's outputs are rated below average. This is mainly due to the low resolution of this method where the VPIs are generated directly from sampling the pixels under each micro-lens. Comparing the totals, the proposed method has a great advantage to generate super resolution VPIs. This is mainly due it being the only one with "Very High" ratings, and due to it also having with greater "high" rating scores than the other methods. The proposed super resolution VPIs are generated from 2DVPIs that are representations of an original scene. Additionally, it was the only tested method able to produce "Excellent" quality 2DVPI representations of a scene with the same large depth of field from each micro-lens, typically called All-In-Focus.



(a)

(b)

Figure 6-19 Comparisons the quality rating values of SR images produced from different algorithms from nine test images, (a) Expert participants /Group 1 and (b) Non-expert/ Group 2.

Table 6-4: Percentage of MOS ratings attributed for overall participants for each 2DVPI generated method.

Method	Below Average			Average & Above			
	Very Low	Low	Total	Medium	High	Very High	Total
Sampled VPI	11.1%	77.8%	87.9%	11.1%	0%	0%	11.1%
Vandewalle et al.	0%	22.2%	22.2%	77.8%	0%	0.%	77.7%
Keran et al.'s	0%	44.4%	44.4%	55.6%	0%	0.%	55.5%
Proposed Algorithm	0%	0%	0%	11.1%	83.3%	5.6%	88.9%

This All-in-Focus production of VPI is an example of the capability of the proposed method to generate SRVPI. It is a novel, fully automated method for generating 2D images, and when implemented into H3DIs system that has micro-images with a uniform structure, the 3D scene relative depth can be estimated with correct localisation and high accuracy. Moreover, the proposed SRVPI solution outperforms all the available alternative methods, where the "poor" condition is almost zero compared to others methods as shown in Figure 6.21.

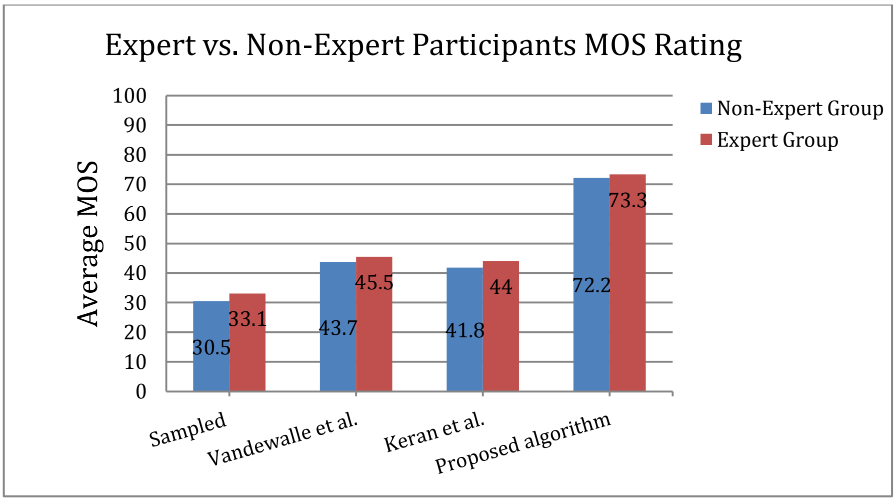


Figure 6-20 Average human score correlation of expert versus non-expert participants for all nine images that generated from different methods.

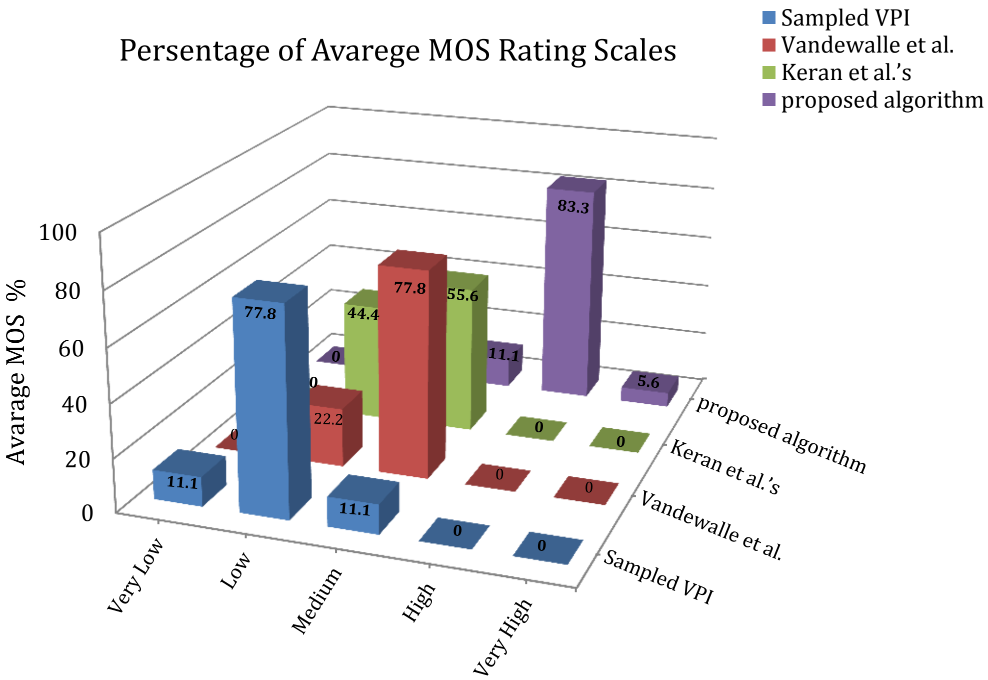


Figure 6-21 Comparisons the percentage of average MOS rating scales for the four methods of all participants.

6.6 Summary

A method to render 3D objects from images so that the information from H3DIs is presented properly was described in this chapter. The work focused on the use of algorithms that gave binary assignments to foreground/background pixels. Therefore the method of 3D object segmentation presented means that the objects in the scene are separated from the background. This work provides a novel method of automatically producing a 3D interactive map (3DIM) that highlights the areas that the viewer is expected to focus on. In the other words, this method finds the objects in the scene that the observer will be interested in seeing. This process incorporates the 2D construction of the object surface plane and the special integration of 3D volume that improves the perspective view. This means that the 3D interactive map can be considered to be a fusing of 2D and 3D techniques. The use of the hybrid technique means that better information and the geometrical shape of the objects in space are obtained.

Two methods were developed to extract the 2D foreground object that enables the 3D interactive map to be developed while giving a robust link between the 2D and 3D views. The methods were based on the principles of simple implementation and full automation so no human input was needed when generating the 3D interactive map. In addition, they also had to be efficient when generating the 3D objects from the H3DI system. With this approach the 3D object segmentation was significantly improved via the use of better determination of foreground/background pixels as well as object boundaries.

The first technique, "morphology-based-threshold" was a fully automatic morphological segmentation algorithm that was used with the reference VPI. A binary foreground mask was produced using the VPI's special intensity distribution and a morphological operation. A threshold value is maintained at a set value and pixels with high-intensity values are those that are important, while the background is composed of low intensity pixels. It was shown that the border was well defined and helped to enhance the object removal from the depth map while also filtering out noise in the image. The evaluation showed that this method reached the required quality level. The 3DIM identified the location of the objects of interest and their precise contours, which enabled their separation from the background. It should be noted that both objective and subjective evaluation of the method showed that small over-segmentation errors occurred where potions of the background were included as part of the foreground object. These were due to false edges of the markers due to the grey level distribution and the SE kernel size, and this was a common problem with all the VPIs used. This means that the ability to define an object that has a very detailed border will significantly be related to the match approximation between the foreground object mask and the object.

So that the over-segmentation problem encountered above could be prevented the second method uses 2D segment regions with closed contours on the VPI. This technique produces a 3DIM that automatically integrates exploitation of the available depth estimation technique as well as a feature edge segmentation method. The information related to the edges comes from a single feature-edge map obtained from Multi-Quantized Local Histogram (MQLH) edge maps, and this is from where the geometrical feature-edges of the object are acquired. 2D foreground masks are produced from the 2D segmentation algorithm that integrates the feature-edge map and region growing techniques. Then the extracted features from the feature-edge map are the basis for the 3D cues map that comes from the VPIs displacement. The validity of the method to extract 3D objects from a scene was determined.

Objective and subjective evaluations showed that the second method to form 3DIMs had better performance than the first method and an external, well known, contour object segmentation technique. The second method produces outputs that are extremely close to the reference object. This means that the results from the 3DIM when this filter is used are of good quality. As high quality feature-edge maps can be produced that are reliant on the trainer VPI's resolution it is possible to obtain high resolution images from a set of low resolution VPIs. The second method is reliant on the calculated edge-map as the object border must be complete for this technique to give successful segmentation. The high quality feature-edges are used in the second method to produce a 3D information map.

The 3DIM creation methods are capable of obtaining a 3D object from an image, but they are more effective if a high resolution EI that has an extended baseline is used to fully identify high accuracy depth cues map and extract the 3D object. It is also required to ignore VPIs that have poor feature correlation so that the 3DIM can be more accurate and the process takes place faster. This means that a more accurate 3D object that has few if any errors will lead to a more accurate 3D information map estimation algorithm.

CHAPTER 7

Conclusions and Future Work

7.1 Summary

The present research has approached the area of 3D technology through the use of a H3DI system. Holoscopy is a 3D technology that has the potential to solve some of the limitations of current 3D technology, such as delivery of the depth perception through special glasses and visual discomfort (eye fatigue). In holoscopic 3D technology, the perception of depth does not require glasses or any other gear whatsoever to aid the illusion. It produces a true sensation of depth by using natural light field reconstruction, projecting light rays in the direction they were travelling at capture time, resulting in an accurate reconstruction of the light at the moment of capture with natural horizontal and vertical parallaxes. The principal aim of this research was to study the subject of delivering true 3D depth, generating high resolution multi-views and extracting 3D objects in the scene from a H3DI system.

This thesis set out to explore methods for the effective use of 3D cues, in the challenging task of delivering a high quality 3D image from a set of low resolution VPIs. The specific work of the author was to investigate a particular technique that would be appropriate for generating 3D depth maps by converting the H3DI into 2D images of objects named VPIs. The benefit of using VPIs was due to the existence of the high correlation between the features of these extracted images. It is however important to point out that the current available display technologies are still not compatible with H3DI formats. Therefore, it was essential to transform the 3D holoscopic formats into 2D formats to present the 3D image formats.

Holoscopy 3D technology has arrived at the professional market and soon will hit the consumer market as a consequence of ambitious projects like the 3DVIVANT PROJECT. It is worth pointing out that this research work was developed as part of the 3D VIVANT project. With the appearance of a company (and its products) called Lytro, a 3D holoscopic camera is available to the public and everyone can produce their own holoscopic content. However, the display technology for H3DIs is not yet at an affordable price for the common buyer, which might be changed in the future by the efforts of researchers. The H3DI technique proved that it is more appealing than other 3D techniques due its ability to create a real 3D image in full colour (true 3D model) and deliver a rich viewing sensation without eye fatigue and free of glasses from a single aperture.

7.2 Conclusions

This thesis set out to explore methods for the effective use of 3D information in the challenging task of investigating techniques that are able to achieve wide generalisation capabilities and with a trade-off between accuracy and speed for estimating precise depth perception. The work was motivated by the prices of the display technology for H3DIs being too much for public buyers.

The first challenge addressed was the lack of generalization in [23, 24] the algorithm related to the estimation of 3D depth from 1D H3DIs, i.e. horizontal UH3DIs. The work was extended for the computationally estimation of full parallax 3D information maps from 2D OH3DIs, taking in to account the generalization capability of the depth-through-disparity estimation algorithm, accuracy and speed. To this end, it was required to modify the multi-baseline algorithm by incorporating an adaptive weighting factor shape window technique to enhance the estimation of the 3D object's position. The adaptive aggregation cost window was used to prevent multiple neighbouring blocks from being identified as interesting in the same feature and for identifying large support areas in untextured zones and near the boundaries of objects. This was achieved by ensuring that extra emphasis was placed on the centre block rather than the surrounding blocks. An obvious improvement over the method proposed by [23, 24] was proven to be effective via experiments on both UH3DIs and OH3DIs. Employing an adaptive aggregation cost window served to force the algorithm to work with complex object scenes in OH3DIs. This lead to a rise in average precision of the depth by focusing the estimation into areas where structural features are most valuable. However, it was noted that, varying the threshold lead to different behaviours of the setting features for each image, and this led to a loss of useful information due to lack of control over the setting and extraction of reliable features. The threshold for the setting of the features was determined manually for each image, which is very time consuming, as there are differences between all the images that could result in false matching and faults in the depth estimation process.

To overcome the above drawback, this work developed and implemented a new corresponding and matching technique that was based on automatic thresholding 3D features. The automatic identification of thresholds highlighted the fact that sparse features can sometimes lead to improved accuracy in the depth in addition to increased computation speed due to the performance generalization of the algorithm. Therefore, the next challenge was to identify algorithms that demonstrate the simplest way that H3DIs can have their depth estimated. Accordingly, the development of automatic initialization techniques for creating and manipulating 3D models from H3DIs was considered. This was the first and crucial step towards determining the depth via the extracted VPIs during the matching of the extracted feature. Where the image is simplified during feature selection, it produces a form where it is easier to analyse and determine the positions of objects.

Two novel automatic local feature (point or edge) setting and extracting techniques were elaborated that were built on binary data. The techniques were simple and widely applicable as well as being easy to compute. The first algorithm was based on the intensity distribution of sample variance of sub-dividing non-overlapping regions of the training VPI. The optimal threshold was the mean value of the sub-dividing regions, which was then used for extracting and setting distinguishing feature-points on the training VPI. The standard feature point's profile that represents a particular feature was generated using the distribution of the optimal threshold along the viewpoint image's patch (block). The feature's descriptors were stored in a binary form. This is because the binary descriptors are less demanding computationally than real value descriptors. Cross-validation was employed to optimize the performance of the estimated feature-points process, and thus led to reducing the computational time in matching and corresponding process. The proposed AFP algorithm outperformed the well-known SIFT and SURF descriptors when setting and extracting the distinguished feature blocks. Using both real-world UH3DI and OH3DI data, satisfactory quality depth maps were produced while the consistency of the measurements were proven in the OH3DI “Box”. However, fine details in the object’s contours were not always obtained with high accuracy. This happens mainly for narrow parts of the objects or small objects due to the dissimilarity between them and the surrounding objects, which are very noisy, even for the human eye in those low resolution viewpoint images. However, even in such cases, a satisfactory approximation of the object contour can be obtained that makes this kind of object perceivable.

Therefore, the other challenge here was to generate a foreground mask to separate the objects in the scene from their background, i.e. surrounding objects or noise, for further enhancement to improve the precision of the object depth map. In other words, to address the problem of representing the extraction of correct information from the H3DI system. The proposed fully automatic morphological segmentation algorithm named “morphology-based-threshold” was employed on the training VPI. The technique was mainly concerned with developing algorithms to show the simplest way of generating the 3D object in terms of determining the binary foreground/background pixel assignments. A morphological operations and spatial intensity distribution was employed on the training reference VPI to generate a 2D binary foreground mask assuming high-intensity pixels are of interest, and low intensity pixels are background. Incorporating the new 2D foreground mask with the estimated 3D cues map produced an interactive map (3DIM) that carries information about where on the image the viewer is expected to look. This interactive map will be encoded together with the visual content to identify the applicable foreground regions and depth information map to recognize 3D objects in the scene. Experimental results show that the contours closely follow the shape of the object, and thus can be used to improve the separation of the object’s depth information map or can also be implemented as a filter to remove noises associated with the images. Accordingly, the 3D objects were correctly localized**.**

However, objective and subjective evaluation criteria identified over-segmentation errors between the created foreground object’s masks and the object, where background portions were attached to the reference segmentation of the object. This mainly occurred due to the false edges of the markers obtained, which highly depends on the grey level distribution (high variation between background and object). Therefore, separating an object with a very detailed border is highly dependent on the degree of match approximation between the created foreground object masks and the object. To improve the 3DIM algorithm performance an automatic edge detection technique was investigated for the setting and detection of feature blocks to identify the object contours and depth information map.

The next challenge was the incorporation of the edge detection technique into the modified multi-baseline disparity algorithm to generate a new form of 3DIM. The previous technique for setting and extracting feature descriptors for depth estimation could not identify 3D objects or 3DIMs with very detailed borders. Thus a new formulation for compatible 3DIM estimation based on robust edge detection features was introduced. The novelty of this work lies in automatically generating a 3DIM that combines exploitation of the available depth estimation technique with the segmentation technique via a single feature-edges map. The feature-edge descriptors were scale invariant, illumination invariant and noise resistant, and thus were considered to be robust and efficient. This generated robust feature-edges information for the shape boundaries of objects that was closely related to the geometry of the 3D scene, and background clutter was completely removed. The extracted feature-edges were exploited in the two aspects of 3DIM production; depth estimation and foreground mask generation for 3D object segmentation. The 3D object map was estimated from an adaptive hybrid depth map estimation technique that eliminates spurious edges in the depth edge map via alternative implementation of feature edge detection. This part considered the contour based depth estimation problem. The 2D foreground segmentation algorithm was generated from the integration of region growth and edge detection techniques to generate a foreground object mask. The technique automatically used the extracted feature-edges as seed pixels and guided the process of region growing by iteration, which adds to these seed points all the pixels that are connected to it. The advantage of this technique is that no homogeneity criterion threshold is required in the region growing process and the seeded region growing starts from all the seeds at the same time.

Experimental results of the new algorithm (AFE)showed it was an improvement as at the nearest depth of 600-1200 mm it gave an error of less than 1.7% against the previous (AFP algorithm) error of 1.9%. The AFE algorithm also showed improvement on the depth map contour and gives a very good depth location. However, there is a slight chance of false matches being generated when the matching of the corresponding feature edge block was undertaken near untextured areas or those with similar texture patterns, such as the ambiguities observed on the flat image "Palm". Regions that are textureless are likely to have the most prominent errors due to the thinning edge term (non-maximum suppression) that causes viable candidates to be suppressed when present in weakly textured regions.

Both depth estimation techniques have had their effectiveness proven when generating accurate depth measurements via mathematical and experimental means. However, objective and subjective segmentation quality evaluations confirm that the performance of the second proposed method to create 3DIMs outperforms the first and a well-known active contour object segmentation technique. This is due to the closed border of the object, which was represented by the feature-edge map. Where, the successful segmentation is highly dependent on the closed border of the object, which has been achieved within this technique. Additionally, this approach was far more applicable to real computer vision tasks than the previously proposed technique, when the same feature-edges are directly employed to generate a foreground object mask. The promising outcomes of the 3DIMs based on the edge detection technique have shown that the 3D object depth segment contour has performed well with only minor errors, despite the slight chance of false matches causing some ambiguities that reduce the accuracy of the depth map. Therefore, enhancement of the VPI's resolution was essential to increase the accurate performance of the 3DIM algorithm and to obtain high depth information maps. This enabled the extracted LRVPIs to produce high-resolution image features and thus improved the depth map's visual quality.

Finally, the super-resolution techniques of image processing were exploited to generate more superior feature-edges for estimating the super 3DIM, thus to increase the visual quality of the 3D object in the scene. A H3DI system was used to generate

SRVPIs from many registered groups of sampled LRVPIs. The new method transforms and refocuses the set of orthographic projection LRVPIs into a perspective projection SRVPI using up-sampling-shift and integration image processing techniques. Due to the sub-pixel shifts process, which introduced the non-redundant information that is present in the set of LRVPIs, the visual quality was improved through the use of a new interpolation approach. This enabled the added high frequency information contained within the scene to be transferred to the final SRVPI. Due to over or under fitting during the reconstruction of the SRVPI associated with fixed shifting of the neighbouring LRVPIs, there was often blurring. Therefore, a simple and effective de-blurring filter technique was investigated to suppress noise while maintaining the SR image's details. This SRVPI was exploited to generate more advanced and reliable feature blocks through the implementation of the second proposed AFE algorithm.

Significantly, there was another novel aspect associated with the creation of SRVPIs. Thus the 3D multi-view technique was proposed, employing sub-pixel shift techniques to create super resolution multi-view images from a H3DI system without complex and expensive multi-camera calibration. The use of multi-view SRVPIs successfully estimated depth information maps with high accuracy. This is due to the rich information associated with the created SRVPIs that contain valuable information, as these results come from large image sensors. Results showed that depth accuracy and a fast execution time have been achieved via the use of the multi-view SRVPIs. This also demonstrated the generalization of the approach to estimation of the depth information map on the two types of H3DIs. In other words, incorporating the super-resolution technique with the technique of H3DI significantly improved the performance of the depth estimation algorithm in the context of the generalization capability, the speed and the quality. Generally, the experimental results of the proposed technique showed excellent ability to measure and determine depth maps with high visual quality from H3DIs.

An investigation to evaluate all the creation phases of the 3DIM technique for segmenting a 3D object in the scene was carried out. This was due to the absence of the reference (no-reference) ground-truth image for the quality assessment metric, where the VPIs are pixels sampled from the EIs. Hence, a subjective quality assessment seems to be promising to assess the fully automated algorithm. This included the quality assessment of the extracted 2D SRVPI and 2D foreground mask as well as the accuracy assessment of the estimated depth map from the H3DIs. Subjective evaluation confirmed the outperformance of the second proposed technique for all 3DIM creation phases in comparison to the first proposed technique and active contour model. It is important to point out that this technique benefited significantly from the outcome of the feature-edge detection algorithm, which aided the setting and extraction of a high quality feature-edge map that in turn was highly dependent upon the resolution of the extracted VPI.

7.3 Suggestions and Future Work

Upon implementation, experimentation and review of the test results, some enhancements were identified for the proposed techniques in the thesis.

7.3.1 Holoscopic 3D Camera Lens Correction

There is not a standard technique for holoscopic 3D camera lens correction, and the proposed technique utilizes VPIs extracted from a H3DI, which includes lens errors such as barrel distortions and unbalance elemental images. There is a need for a robust holoscopic 3D camera lens correction algorithm to minimize the noise.

7.3.2 Generalisation Capabilities

In this approach, a special focus was given to tasks on the principle of invariability algorithms, i.e. wide generalisation capabilities (no restraints were placed on the algorithms). These have the ability to measure and estimate 3D depth from both types of H3DIs, real and CG images. Therefore, with respect to the generalisation capability algorithms, the comprehensive experiments presented in this approach clearly result in increases in the error score depth estimation when the algorithm is forced to work in various acquisition environments. In this regard, there is still significant room for the development of the best choices for regularization of the parameters automatically.

7.3.3 2D Viewpoint Image Resolution

The obtained SRVPIs were focused at a particular depth causing blurring effects where the object depth was out of plane. This is due to the choice of a fixed shift value, i.e. one pixel shifting. In other words, the choice of one shift value returns one depth plane ‘in-focus’. Experimentally, there was still some blurriness available even after implementing the de-blurring filter. Thus, a different shift value would correspond to a different depth plane. Hence, employing the Michelson contrast algorithm on different pitch sized windows to estimate the blurring noise in all the depth plane images can be used to return all-in-focus images and will increase the efficiency of the super resolution process from LRVPs.

7.3.4 4D Modelling

To date, 3D depth estimation is performed on spatial information from the H3DI; however, there are sufficient low level cues in the H3DI to reconstruct a 3D model of a real scene. There is a need for a 4D modelling algorithm that utilizes holoscopic low level cues to reconstruct a 3D model of a real scene captured by a H3DI camera.

7.3.5 Holoscopic Elemental Image Resolution Enhancement

The research exploited improvements in the VPI resolution, which resulted in a great difference in the 3D depth estimation algorithm. However, due to limited CCD/CMOS resolution, the H3DI still has low resolution elemental images, and now the next step is to investigate the enhancement of the elemental image resolution computationally in post-production.

7.3.6 Computational Efficiency

Currently, the proposed algorithms have been implemented in MATLAB. It would be possible to improve the computational efficiency to speed the processing if developing the programs in C++ for the Central Processing Unit (CPU) as well as for the Graphics Processing Unit (GPU). The use of the GPU gives immensely superior graphics processing power that allows for fast processing.

7.3.7 H3DI Quality Matrices

For the successful penetration of 3D technology to the market, it is essential to ensure that the new experience is superior to the currently available technique to the users. Therefore, the evaluation of 3D quality is considered as an enormous challenge. Here, new investigation techniques are needed to assess this kind of experience. The current subjective standards tests are not fully dedicated to the evaluation of 3D content quality of H3DI system scenarios. This is due to the non-availability of reference/ground truth image from H3DI system. Therefore, more investigation and research are necessary for developing a 3D quality measure metric and further studies will require for subjective assessment methodologies.

REFERENCES

[1] N. A. Dodgson, “Autosteroscopic 3D Display,” Published by the IEEE Computer Society, pp. 31-36, Aug. 2005.

[2] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, “Three-Dimensional Display Technologies of Recent Interest: Principles, Status, and Issues,” Applied optics, Vol. 50, No. 34, pp. 87–115, Dec. 2011.

[3] S. Zinger, D. Ruijters, and P. H. N. de With, “iGLANCE Project: Free-Viewpoint 3D Video,” 17th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG), 2009.

[4] D. Ruijters, “iGLANCE: Transmission to Medical High Definition Autostereoscopic Displays,” 3DTV Conference: The True Vision Capture, Transmission and Display of 3D Video, 2009.

[5] L. Onural, “Television in 3-D: What are the Prospects,” Proc. IEEE, Vol. 95, No. 6, 2007.

[6] http://www.codex99.com/photography/4.html. [Online Accessed: Jan. 2014].

[7] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. V. Kopylow, “A Survey of 3DTV Displays: Techniques and Technologies,” IEEE Trans. Circuits Syst. Video Technol., Vol. 17, No. 11, pp. 1647–1658, Nov. 2007.

[8] H. M. Ozaktas and L. Onural, Eds., “Three-Dimensional Television: Capture, Transmission, Display,” Heidelberg, Germany: Springer, ISBN: 78-3-540-72531-2, 2007.

[9] N. A. Valyus, Stereoscopy, 1st Edition. London: Focal Press, 1966.

[10] S. Monaleche, A. Aggoun, M. McCormick, N. Davies, S.Y. Kung, “Analytical Model of a 3D Recording Camera System Using Circular and Hexagonal Based Spherical Microlenses,” Journal of Optical Society America A, Vol. 18, No. 8, pp. 1814-1821, 2001.

[11] R. B. A. Tanjung, X. Xu, X. Liang, S. Solanki, Y. Pan, F. Farbiz, B. Xu, and T.-C. Chong, "Digital Holographic Three-Dimensional Display of 50-Mpixel Holograms Using a Two-Axis Scanning Mirror Device," *Opt. Eng.*, Vol. 49, No. 2, p. 025801-2, Mar. 2010.

[12] B. G. Blundell, A. J. Schwarz, and D. K. Horrell, "Volumetric Three Dimensional Display Systems: Their Past, Present and Future," *Eng. Sci. Edu. J.*, Vol. 2, No. 5, pp. 196–200, Oct. 1993.

[13] Y. Zhang, Q. Ji, and W. Zhang, "Multi-View Autostereoscopic 3D Display," in Int. Conf. on Opt. Photon. Energy Eng. (OPEE), 2010.

[14] S. Manolache, S.Y. Kung, M. McCormick, and A. Aggoun, "3D-Object Space Reconstruction From Planar Recorded Data of 3D-Integral Images," Journal of VLSI Signal Processing Systems, Vol. 35, pp. 5-18, 2003.

[15] N. A. Dodgson, "Analysis of the Viewing Zone of the Cambridge Autostereoscopic Display," Applied Optics, Vol. 35, No. 10, pp. 1705-1710, 1996.

[16] P. Harmann, "Retro reflective Screens and Their Application to Autostereoscopic Displays," Proc. SPIE—Int'l Soc. for Optical Eng., Vol. 3012, pp. 145-153, 1997.

[17] A. Aggoun, "3D Holoscopic Imaging Technology for Real-Time Volume Processing and Display," High-Quality Visual Experience Signals and Communication Technology, IV, pp. 411-428, 2010.

[18] M. G. Lippmann, "La Photographie Integrale," Comptes-Rendus de l'Academie des Sciences, Vol. 146, pp. 446-551, 1908.

[19] A. BOGUSZ, "Holoscopy and Holographic Principles," Journal Optics (Paris), Vol. 20, No. 6, pp. 281-284, 1989.

[20] G. Lawton, ‘‘3D Displays Without Glasses: Coming to a Screen Near Your Computer,'' Computer, Vol. 44, No. 1, pp. 17-19, 2011.

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

[22] Y. Kim, K. Hong, and B. Lee, "Recent Researches Based On Integral Imaging Display Method," 3D Research, Vol. 1, No. 1, pp. 17–27, Aug. 2011.

[23] C. H. Wu, M. McCormick, A. Aggoun, and S.Y. Kung, "Depth Mapping of Integral Images Through Viewpoints Image Extraction With a Haybrid Disparity Analysis Algorithm," Journal of Display Technology, Vol. 4, No. 1, pp. 101-108, 2008.

[24] H.-x. Wang, Z.-li Xu, Z.-p., and C. H. Wu, "3D Reconstruction from Integral Images based on Interpolation Algorithm," Proc. SPIE 7655, 5th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Advanced Optical Manufacturing Technologies, Vol. 7655, pp. 1-7, Oct. 2010.

[25] W. Ijsselsteijn, H. de Ridder, and J. Vliegen, "Effects of Stereoscopic Filming Parameters and Display Duration on The Subjective Assessment of Eye Strain," In *Proc. SPIE Stereosc. Displays and Virtual Reality Syst. VII*, Vol. 3957, pp. 12–22, 2000.

[26] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Integral Imaging with Improved Depth of Field by Use of Amplitude Modulated Microlens Array," Appl. Opt., Vol. 43, pp. 5806–5813, 2004.

[27] J.-S. Jang and B. Javidi, "Improved Viewing Resolution of Three-Dimensional Integral Imaging by Use of Nonstationary Micro-Optics," *Opt. Lett.*, Vol. 27, pp. 324–326, 2002.

[28] M. Ye et al., "A Survey on Human Motion Analysis from Depth Data," Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications Lecture Notes in Computer Science, Vol. 8200, pp. 149-187, 2013.

[29] K. Kulhavy, "Image Analysis of Mona Lisa by the Human Visual System, up to Simple Cells in Visual Cortex," July, 2012. http://commons.wikimedia.org/w /index.php?title=File:Lisa_analysis.png. [Online Accessed: Jan. 2014]

[30] S. R. Barry, "Fixing My Gaze," A Scientist's Journey into Seeing in Three Dimensions, Basic Books, pp. 101-120, 2009.

[31] S. Pastoor, "Human Factors of 3D Imaging: Results of Recent Research At Heinrich-Hertz-Institute Berlin," Proc. 2nd Int. Display Workshop, Hamamatsu, pp. 69-72, 1995.

[32] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-Dimensional Display Technologies of Recent Interest: Principles, Status, And Issues," Applied optics, Vol. 50, No. 34, pp. H87-115, Dec-2011.

[33] L. Lipton, Foundations of the Stereoscopic Cinema, Van Nostrand Reinhold, 1982; www.stereoscopic.org. [Online Accessed: Feb. 2014]

[34] M. Wöpking, "Viewing Comfort with Stereoscopic Pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," Journal of the SID, Vol. 3, No. 3, pp. 101-103, 1995.

[35] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in Stereoscopic and Autostereoscopic Displays," Proceedings of the IEEE, Vol. 99, No. 4, pp. 540-555, Apr. 2011.

[36] http://www.bekkahwalker.net/comt111a/websites_11/camacho_site/gallery.html.[Online Accessed: Jan. 2014].

[37] A. K. Srivastava, de Bougrenet de la Tocnaye, J. L. and Dupont, L., "Liquid Crystal Active Glasses for 3D Cinema," IEEE, Vol. 6, Issue:10, pp. 522 - 530, 2010.

[38] E. Dubois "A projection Method to Generate Anaglyph Stereo Images," Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP'01). IEEE International Conference, Vol. 3, pp. 1661-1664, 2001.

[39] M. Brian, "How 3-D Glasses Work," http://science.howstuffworks.com/3-d-glasses2.htm. [Online Accessed: Jan. 2014]

[40] D. Getty, "Stereo Imaging and Display: Review of Stereo Vision," http://siim.org/books/displays/chapter-4-stereo-imaging-displays#consequences. [Online Accessed: Jan. 2014]

[41] S. Li, L. Ma, and K. N. Ngan, "Anaglyph Image Generation by Matching Color Appearance Attributes," Signal Processing: Image Communica-tion, Vol. 28, pp. 597-607, 2013.

[42] S. Omeltshenko,"3-D-display", IEEE Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), International Conference on, pp. 108, Feb. 2010.

[43] http://3dcinecast.blogspot.co.uk/2010/11/shift-to-3d-reignites-tv-image-quality.html. [Online Accessed: Jan. 2014]

[44] FOS series, LC-Tec Displays AB, FOS series model overview, September, 2013, http://www.lc-tec.com/UserFiles/Products/FOS_Specifications/FOS_FOS-AR_specifications_1309.pdf. [Online Accessed: Jan. 2014]

[45] P. Boher, T. Leroux, T. Bignon and V. Collomb-Patton "Multispectral Polarization Viewing Angle Analysis of Circular Polarized Stereoscopic 3D Displays", Proc. SPIE, Vol. 7524, pp. 1-12, 2010.

[46] http://www.maximumpc.com/article/features/white_paper_stereoscopic_imaging. [Online Accessed: Jan. 2014]

[47] A. W. Divelbiss, D. C. Swift and W. V. Tserkovnyuk "3D Stereoscopic Shutter Glass System," 2004, http://www.google.co.uk /patents/ US6727867. [Online Accessed: Jan. 2014]

[48] G. E. Favalora, "Volumetric 3D Displays and Application Infrastructure," IEEE Computer, Vol. 38, pp. 37-44. 2005.

[49] D. Gabor. "Microscopy by reconstructed Wavefronts", Proc. Phys. Soc., No. A194, pp. 454-487, 1949.

[50] http://en.wikipedia.org/wiki/Dennis_Gabor. Available: [Online Accessed: Feb. 2014].

[51] B. Dodson, "New Technology from MIT may enable Cheap, Color, Holographic Video Displays," http://www.gizm ag.com/holograph-3d-color-video-display-inexpensive-mit/28029/pictures#5. [Online Accessed: Jan. 2014]

[52] P. A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W. Y. Hsieh, and M. Kathaperumal, "Holographic Three-Dimensional Telepresence Using Large-Area Photorefractive Polymer," Nature, Vol. 468, pp. 80-83. 2010.

[53] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, " Three-Dimensional Displays: A Review and Applications Analysis," IEEE Transactions on Broadcasting, Vol. 57, No. 2, pp. 362-371, 2011.

[54] J. Andrew, M. D. Ian, Y. Hideshi, B. Mark, and D. Paul, "Rendering for an Interactive 360 Light Field Display," SIGGRAPH Papers Proceedings, pp. 34-39, 2007.

[55] http://www.youtube.com/watch?v=eNWJ9XtRhLw. [Online Accessed: Jan. 2014]

[56] N.A. Dodgson et al., "A 50-Inch Time-Multiplexed Autostereoscopic Display," Proc. SPIE-Int'l Soc. for Optical Eng., Vol. 3957, pp. 177-183. 2001.

[57] A. Boev, R. Bregovic, and A. Gotchev, " Signal Processing for Steroscopic and Multi-view 3D Display," Chapter in Handbook of signal processing systems, 2nd edition, edited by Bhattacharyya, E. Deprettere, R. Leupers, and J. Takala, Springer, pp. 3-37, 2013.

[58] N. A. Dodgson, "Autostereoscopic 3D Displays," published by the IEEE computer society, pp. 31-36, Aug. 2005.

[59] C. Riechert, F. Zilly, P. Kauff, J. Güther, and R. Schäfer, "Fully Automatic Stereo to Multi-view Conversion in Autostereoscopic Display," International Broadcasting Convention, pp. 8-14, Sep. 2012.

[60] N. A. Dodgson, "Multi-view Autostereoscopic 3D Display," Stanford Workshop in 3D Imaging, Stanford, pp. 1-42, Jan. 2011.

[61] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. R. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High Performance Imaging using Large Camera Arrays," ACM Trans. Graph., Vol. 24, No. 3, 2005.

[62] A. Davis, M. Levoy, and F. Durand, "Unstructured light fields," Comput. Graph. Forum 31, pp. 305-314, 2012.

[63] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, R Raskar "Polarization Fields: Dynamic Light Field Display using Multi-layer LCDs. ACM Trans. Graph. (SIGGRAPH Asia), Vol. 30, pp. 1-9, 2011.

[64] 3D VIVANT-Team, "HoloVisio 3D Display," http://www.holografika.com/. [Online Accessed: Jan. 2014]

[65] T. Peterka, R. L. Kooima, D. J. Sandin, A. Johnson, J. Leigh, and T. A. DeFanti, "Advances in the Dynallax Solid-State Dynamic Parallax Barrier Autostereoscopic Visualization Display System," IEEE Trans. Vis. Comput. Graph., Vol. 14, No. 3, pp. 487-499, 2008.

[66] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-Adaptive Parallax Barriers: Optimizing Dual-Layer 3D Displays Using Low-Rank Light Field Factorization," ACM Trans. Graph., Vol. 29, No. 6, pp. 163-172, 2010.

[67] H. Stolle, J.-C. Olaya, S. Buschbeck, H. Sahm, and A. Schwerdtner, "Technical Solutions for a Full-resolution Autostereoscopic 2D/3D Display Technology," In Proceedings of SPIE, San Jose, California, Vol. 6803, Feb. 2008.

[68] A. N. Putilin, A. A. Lukianitsa and K. Kanashin, "Stereo Display with Neural Network Image Processing," In SPIE Advanced Display Technologies. Vol. 4511, pp. 245-250, 2001.

[69] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting," ACM Trans. Graph. (SIGGRAPH), Vol. 31, pp. 1-11, 2012.

[70] A. Maimone, G. Wetztein, M. Hirsch, D. Lanman, R. Raskar and N. Fuchs, "Focus 3D: Compressive Accommodation Display," ACM Transactions on Graphics (TOG), New York, NY, USA , Article 153, Vol. 32, No. 5, pp. 153-165, Sep. 2013.

[71] http://web.media.mit.edu/~gordonw/Focus3D/.[Online Accessed: Feb. 2014].

[72] A. Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L. D. Soares, "Immersive 3D Holoscopic Video System," MultiMedia, IEEE, Vol. 20, No. 1, pp. 28-37, 2013.

[73] G. Lippmann, "Photographies Integrals," ComptesRendus de l'Academic des Sciences, Vol. 146, pp. 446-451, 1908

[74] A. Aggoun, "3D Holoscopic Imaging Technology for Real-Time Volume Processing and Display," In High-Quality Visual Experience, M. Mrak, Marta and Grgic, Mislav and Kunt, Ed. Springer Berlin Heidelberg, pp. 411-428, 2010.

[75] A. P. Sokolov, "Autostereoscopy and Integral Photography by Professor Lippmann's Method," Izd. MGU, Moscow State Univ. Press, 1911.

[76] H. E. Ives, "Optical Properties of a Lippmann Lenticulated Sheet," J. Opt. Soc. Amer., Vol. 21, pp. 171-176, Mar. 1931.

[77] N. Davies and M. McCormick, "Holoscopic Imaging with true 3-D Content in Full Natural Color," The journal of photographic science, Vol. 40, pp. 46-49, 1992.

[78] J-Y Jany, S-H Park, S. Cha, and S-H Shin, "Three-dimensional Integral Imaging for Orthoscopic Real Image Reconstruction," Proceedings of SPIE, Bellingham, WA, Vol. 5636, pp. 379-386, 2005.

[79] A. Chutjian, and R.J. Collier, "Recording and Reconstructing Three-Dimensional Images of Computer Generated Subjects by Lippmann Integral Photography," Appl. Opt. Vol. 7, No. 1, Jan. 1968.

[80] I. Villums, "Optical Imaging System Using Optical Tone-Plate Elements," US Patent 4,878,735, 1989.

[81] M. Hain, W. V. Spiegel, M. Schmiedchen, T. Tschudi, and B. Javidi,"3-D Integral Imaging using Diffractive Fresnel Lens Arrays," Optics Express, Vol. 13, No. 1, pp. 315-326, Jan. 2005.

[82] N. Davis, M. McCormick, and L. Yang, "Three-dimensional Imaging Systems: A new Development," Appl. Opt. Vol. 27, No.21, Nov. 1988.

[83] N. Davis and M. McCormick, "Producing Visual Images," US Patent, 6,614,552, 2003.

[84] Y. Kim, J.-H. Park, H. Choi, S. Jung, S-W Min, and B. Lee, "Viewing-angle-enhanced Integral Imaging System using a Curved Lens Array," Optics Express Journal, Vol. 12, No. 3, pp. 422-429, 2004.

[85] J.-H. Park, J. Kim, Y. Kim, and B. Lee, "Resolution-enhanced Three-dimension/Two-dimension Convertible Display Based on Integral Imaging," Optics Express, Vol. 13, No. 6, pp. 1875-1884, 2005.

[86] R. Martinez-Cuenca, G. Saavedra, M. Martinez-Corral, and B. Javidi, "Progress in 3-D Multi-perspective display by integral imaging," Proc. IEEE , Vol. 97, No. 6, pp. 1067-1077, 2009.

[87] 3DVIVANT, "Deliverable 7.1-System Integration," Version2, http://dea.brunel .ac.uk/3dvivant/assets/documents/WP7%203DVIVANT%20D7.1.pdf, 2013. [Online Accessed: Feb. 2014]

[88] http://search.newport.com/?x2=sku&q2=MALS13. [Online Accessed: Sep. 2014]

[89] http://web.media.mit.edu/~halazar/autostereo/autostereo.html. [Online Accessed: Sep. 2014]

[90] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Multifacet Structure of Observed Reconstructed Integral Images," J. Opt. Soc. Am., Vol. A 22, pp. 597-603, 2005.

[91] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision (2nd ed.)," Published in the United States of America by Cambridge University Press, New York, Apr. 2004.

[92] M. Agrawal, K. Konolige, and R. Bolles, "Localization and Mapping for Autonomous Navigation in Outdoor Terrains: A stereo Vision Approach," In IEEE Workshop on Applications of Computer Vision. Austin, Texas, USA, 2007.

[93] L. Zhang, B. Curless, and S. M. Seitz, "Space time Stereo: Shape Recovery for Dynamic Scenes," In CVPR, Vol. 2, pp. 367-374, 2003.

[94] D. Scharstein and R. Szeliski, "A taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," International Journal of Computer Vision, Vol. 47, No. 1, pp. 7-42, 2002.

[95] L. Nalpantidis, G. C. Sirakoulis, and, A. Gasteratos, "Review of Stereo Vision Algorithms: From Software to Hardware," International Journal of Optomechatronics, Vol. 2, No. 4, pp. 435-462, 2008.

[96] L. Nalpantidis and A. Gasteratos, "Stereo Vision Depth Estimation Methods for Robotic Application," In Depth Map and 3D Imaging Applications: Algorithms and Technologies, Ch. 21, pp 397-417, Aamir Saeed Malik, Tae-Sun Choi and Humaira Nisar (Eds), IGI Global, Hershey PA, USA, 2012.

[97] D. A. Ross, D. Tarlow, and R. S. Zemel, "Learning Articulated Structure and Motion," Int J Comput Vis, Vol. 88, pp. 214-237, 2010.

[98] A. Hosni, C. Rhemann, M. Bleyer, and M. Gelautz, "Temporally Consistent Disparity and Optical Flow Via Efficient Spatio-Temporal Filtering," PSIVT, Part I, LNCS 7087, pp. 165-177, 2011.

[99] C. C. Pham, V. D. Nguyen and, J. W. Jeon, "Efficient Spatio-Temporal Local Stereo Matching using Information Permeability Filtering," IEEE on ICIP, pp. 2965-2968, 2012.

[100] A. Castro, Y. Frauel, and B. Javidi, "Integral Imaging with Large Depth of Field using An Asymmetric Phase Mask," Opt. Express, Vol. 15, pp. 10266-10273, 2007.

[101] Y. Kim, J.-H. Park, S.-W. Min, S. Jung, H. Choi, and B. Lee, "Wide Viewing-Angle Integral Three Dimensional Imaging System by Curving A Screen and A Lens Array," Appl. Opt., Vol. 44, pp. 546-552, 2005.

[102] M. Miura, J. Arai, T. Mishina, M. Okui, and F. Okano, "Integral Imaging System with Enlarged Horizontal Viewing Angle," Proc. SPIE 8384, pp. 838-40O, 2012.

[103] H. Navarro, R. Martínez-Cuenca, G. Saavedra, M. Martínez Corral, and B. Javidi, "3D Integral Imaging Display by Smart Pseudoscopic-to-Orthoscopic Conversion (SPOC)," Opt. Express, Vol. 18, pp. 25573-25583, 2010.

[104] Y. Piao, Y. Wang and J. Zang, "Computational Integral Imaging Reconstruction Technique with High Image Resolution," IEEE, Computer Society, Asia-Pacific Conference on Information Processing, pp. 160-163, 2009.

[105] L. Onural, "Television in 3-D: What are the Prospects," Proc. IEEE, Vol.95, No. 6, pp.1143-1145, 2007.

[106] S. Manolache, M. McCormick, and S. Y. Kung, "Hierarchical Adaptive Regularization Method for Depth Extraction from Planar Recording of 3-D-Integral Images," In Proc. ICASSP, Vol. 3, pp. 1433-1436, 2001.

[107] M. Bertero, and P. Boccacci, "Inverse Problems and Computational Imaging," in Encyclopedia of Modern Optics, eds. R. D. Guenther, D. G. Steel, and L. Bayvel, Vol. 2, pp. 118-127, 2004.

[108] S. Manolache and S. Y. Kung, A. Aggoun, and M. McCormick "3-Dobject Space Reconstruction from Planar Recorded Data of 3-D-Integral Images," J. VLSI Signal Process. Syst. for Signal, Image, and Video Technol., Vol. 35, No. 5, pp. 18-35, 2003.

[109] S. Cirstea, A. Aggoun, and M. McCormick, "Depth Extraction from 3D Integral Images Approached as an Inverse Problem," Proc. IEEE Int. Symposium on Industrial Electronics, Cambridge, UK, pp. 798-802, 2008.

[110] L. Landweber, "An Iteration Formula for Fredholm Integral Equations of the First Kind," Amer. J. Math., Vol. 73, pp. 615-624, 1951.

[111] A.N. Tikhonov and V.Y. Arsenin, "Solutions of Ill-Posed Problems," Wiley, New York, 1977.

[112] C. Wu, A. Aggoun, M. McCormick, and S. Y. Kung, "Depth Extraction From Unidirectional Integral Image Using A Modified Multi-Baseline Technique," Proc. SPIE, Vol. 4660, pp. 135-143, 2002.

[113] J.-H. Park, S. Jung, H. Choi, and B. Lee, "A Novel Depth Extraction Algorithm Incorporating a Lens Array and a Camera by Reassembling Pixel Columns of Elemental Images," Conference on Optical Information Processing Technology, SPIE Photonics Asia, Proc. SPIE, Vol. 4929, Shanghai, China, pp. 49-58, Oct.-2002.

[114] C. H. Wu, M. McCormick, A. Aggoun, and S.-Y. Kung, “Depth map from Unidirectional Integral Images using a Disparity Algorithm based on

Neighbourhood Constraint and Relaxation," IET, International Conference on Visual Information Engineering, pp. 65-68, 2003.

[115] D. Zarpalas, I. Biperis, E. Fotiadou, E. Lyka, P. Daras, M. G. Strintzis, "Depth Estimation In Integral Images by Anchoring Optimization Techniques," ICME, IEEE International Conference on Multimedia and Expo, pp.1-6, 2011.

[116] D. Zarpalas et al., "Anchoring-Graph-Cuts towards Accurate Depth Estimation in Integral Images," IEEE J. Display Technology, Vol. 8, No. 7, pp. 405-417, 2012.

[117] H. Yoo, "Artifact Analysis and Image Enhancement in Three-Dimensional Computational Integral Imaging Using Smooth Windowing Technique," Opt. Lett, Vol. 36, pp. 2107-2109, 2011.

[118] H. Yoo, "Depth Extraction for 3D Objects Via Windowing Technique In Computational Integral Imaging with A Lenslet Array," Optics and Lasers in Engineering, Vol. 51, pp. 912-915, 2013.

[119] G. Baasantseren, J. Park, N. Kim, and K. Kwon, "Computational Integral Imaging with Enhanced Depth Sensitivity," Journal of Information Display, ISSN, Vol. 10, No. 1, pp.1598-0316, Mar. 2009.

[120] D. Sun, S. Roth, and M. J. Black, "Secrets of Optical Flow Estimation And Their Principles," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2432-2439, 2010.

[121] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," In Proceedings of IEEE Conference on International Conference on Computer Vision, pp. 1-8, 2007.

[122] J. Jung, K. Hong, G. Park, I. Chung, J. Park, and B. Lee, " Reconstruction of Three-Dimensional Occluded Object Using Optical Flow and Triangular Mesh Reconstruction in Integral Imaging," Opt. Express, Vol. 18, pp. 26373-26387, 2010.

[123] L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: From software to hardare," International Journal of Optomechatronics, Vol. 2, No. 4, pp. 435–462, 2008.

[124] L. Nalpantidis and A. Gasteratos, "Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence," Robotics and Autonomous Systems, Vol. 58, pp. 457–464, 2010.

[125] C. H. Wu, M. McCormick, A. Aggoun, and S.-Y. Kung, "Depth measurement from integral images through viewpoint image extraction and a modified multi-baseline disparity analysis algorithm," SPIE, Journal of Electronic Imaging, Vol. 14, No. 2, pp. 1-9, Apr. 2005.

[126] M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No.4, pp. 353-363, 1993.

[127] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory And Experiment," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 16, No. 9, pp. 920–932, 1994.

[128] KJ Yoon, IS IS Kweon, "Adaptive Support-Weight Approach For Correspondence Search," IEEE Trans Pattern Anal Mach Intell , Vol. 28, No. 4, pp. 650-656, 2006.

[129] NYC Chang, TH Tsai, BH Hsu, YC, Chen, TS Chang, "Algorithm and Architecture of Disparity Estimation with Mini-Census Adaptive Support Weight," IEEE Trans Circ Syst Video Technol, Vol. 20, No. 6, pp.792-805, 2010.

[130] M. Gong, R. Yang, L. Wang, "A performance Study on Different Cost Aggregation Approaches Used in Real-Time Stereo Matching," Int J Comput Vis, Vol. 75, No. 2, pp. 283-296, 2007.

[131] G. Jianping, et al., "A New Distance-weighted k-nearest Neighbor Classifier.," Information & Computational Science, pp. 1429-1436, 2012.

[132] ftp://ftp.iti.gr/pub/Holoscopy/Integral_Database/Database/Data/. [Online Accessed: Nov. 2012].

[133] C. Kim and P. Milanfar, "Visual Saliency in Noisy Images," Journal of Vision, Vol. 13, No. 4, pp. 1-14, 2013.

[134] C.-H. Yoo, H.-H. Kang, and E.-S. Kim, "Enhanced Compression of Integral Images by Combined Use of Residual Images and MPEG-4 Algorithm in Three-Dimensional Integral Imaging," Opt. Commun. Vol. 284, No. 20, pp. 4884–4893, 2011.

[135] W. van der Mark, and D.M. Gavrila, "Real-Time Dense Stereo for Intelligent Vehicles," Intelligent Transportation Systems, IEEE, pp.38-50, 2006.

[136] R. Kouskouridas and A. Gasteratos, “Location Assignment of Recognized Objects via a Multi-Camera System,” International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No. 3, pp. 1-18, Sep. 2011.

[137] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary Robust Invariant Scalable Keypoints," in Computer Vision (ICCV), IEEE International Conference, pp. 2548-2555, 2011.

[138] Z. Wang, B. Fan, and F. Wu, "Local Intensity Order Pattern for Feature Description," in Computer Vision (ICCV), IEEE International Conference, pp. 603-610, 2011.

[139] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An Efficient Alternative to Sift or Surf," in Computer Vision (ICCV), IEEE International Conference, pp. 2564-2571, 2011.

[140] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast Retina Key-point," in IEEE Conf. on Computer Vision and Pattern Recognition, pp. 510-517, Jun.2012.

[141] O. Miksik and K. Mikolajczyk, "Evaluation of Local Detectors and Descriptors for Fast Feature Matching," in Pattern Recognition (ICPR), 21st International Conference on. IEEE, pp. 2681-2684, 2012.

[142] S. Qi, L. Zhaorong, J. Jiaya, and T. Chi-Keung, "Fast Image/Video Up Sampling," ACM Transactions on Graphics, Vol. 27, No. 5, Article 153, pp. 1-7, 2008.

[143] D. Lowe, "Distinctive Image Features from Scale-Invariant Key-points," International journal of computer vision, Vol. 60, No. 2, pp. 91-110, 2004.

[144] Y. Bengio and N. Chapados, "Extensions to Metric-Based Model Selection," JMLR, Vol. 3, pp. 1209-1227, 2003.

[145] T.-C. Lin, "A Novel Decision Based Median-Type Filter Using SVM For Image Denoising," International Journal of Innovative, Computing, Information and Control, Vol. 8, No. 5(A), May 2012.

[146] T.-C. Lin and C.-M. Lin, "Decision Based Low-Supper-Middle Filter for Image Processing," International Journal of Innovative, Computing, Information and Control, Vol. 7, No. 10, pp. 5977-5990, 2011.

[147] D. Lowe, "Distinctive Image Features from Scale-invariant Key-points," IJCV, Vol. 60, No. 2, pp. 1-28, 2004.

[148] http://www.vlfeat.org/overview/sift.html. [Online Accessed: April. 2014].

[149] H. Bay, A. Ess , T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding, ELSEVIER, Vol. 110, pp. 346-359, 2007.

[150] http://www.mathworks.co.uk/help/vision/ref/detectsurffeatures.html. [Online Accessed: April. 2014].

[151] D. Marr and E. Hildreth, "Theory of Edge Detection," Proceedings of the Royal Society of London. Series B, Biological Sciences, Vol. 207, No. 1167, pp. 187-217, 1980.

[152] D. R. Waghule and R. S. Ochawar, "Overview on Edge Detection Methods," IEEE, International Conference on Electronic Systems, Signal Processing and Computing Technologies, pp. 151-155, 2014.

[153] W. Liang, C. Da and L. Dun, "A single Threshold Edge Detection Algorithm Based on Multi Scale Fusion," Computational Intelligence and Security, Guangzhou, Vol. 2, pp. 1710-1715, Nov. 2006.

[154] S. Jansi and P. Subashini, "Optimized Adaptive Thresholding Based Edge Detection Method for MRI Brain Images," International Journal of Computer Applications, Published by Foundation of Computer Science, New York, USA, Vol. 51, No. 20, pp. 1-8, Aug. 2012.

[155] M. Khallil and A. Aggoun, "Edge Detection Using Adaptive Local Histogram Analysis," IEEE, ICASSP, pp. 757-760, 2006.

[156] M. Khallil, A. Aggoun, and A. El-mabrouk " Edge detector using local histogram analysis," Proc. SPIE 5150, Visual Communications and Image Processing. Jun. 16, 2003.

[157] A. Aggoun and M. Khallil, "Multi-Resolution Local Histogram Analysis for Edge Detection," IEEE, ICIP, pp. 45-48, 2007.

[158] Q. Lia, J. Yeb and C. Kambhamettu, "Interest Point Detection Using Imbalance Oriented Selection," Pattern Recognition Society. Published by Elsevier Ltd., Vol. 41, pp. 672- 688, 2008.

[159] C. Schmid, R. Mohr and, C. Bauckhage, "Evaluation of Interest Point Detectors," Int. J. Comput. Vis., Vol. 37, No. 2, pp. 151-172, 2000.

[160] K. A. Parulski, L. J. D'Luna, B. L. Benamati, and P. R. Shelley, "High performance digital color video camera," J. Electron. Imaging, Vol. 1, pp. 35–45, 1992.

[161] S. Chaudhuri, Ed., "Super-Resolution Imaging," Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[162] S. Borman and Robert Stevenson, "Spatial Resolution Enhancement of Low Resolution Image Sequences: a comprehensive review with directions for future research," 1998. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33. 5918. [Online Accessed: May. 2014]

[163] R. Sudheer Babu and K.E.Sreenivasa Murthy, "A Survey on the Methods of Super-Resolution Image Reconstruction," International Journal of Computer Applications Vol. 15, No. 2, pp. 1-6, Feb. 2011.

[164] S. Shafique Q., X. M. Li, and T. Ahmad, "Investigating Image Super Resolution Techniques: What to Choose?," ICACT, pp. 642-647, 2012.

[165] J. Tian and K. K. Ma, "A survey on Super-Resolution Imaging," Springer-Verlag London Limited, pp. 1-14, 2011.

[166] R. Ng, M. Levoy, M. Br´edif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a hand-held Plenoptic Camera," Technical Report CSTR, Stanford University, pp. 1-11, Apr. 2005.

[167] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offs through a Bayesian Analysis of Light Field Projections," In European Conference on Computer Vision, No. 14, pp. 619–624, 2008.

[168] R. Keys, "Cubic Convolution Interpolation for Digital Image Processing," IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 29, pp. 1153-1160, December 1981.

[169] H.S. Hou and H.C. Andrews, "Cubic Splines for Image Interpolation and Digital Filtering," IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 26, pp. 508-517, 1978.

[170] H. Ur and D. Gross, "Improved resolution from Sub-pixel Shifted Pictures," CVGIP: Graphical Models and Image Processing, Vol. 54, No. 2, pp. 181-186, 1992.

[171] R. Y. Tsai and T. S. Huang, “Multiple-frame Image Restoration and Registration. In Advances in Computer Vision and Image Processing,” Greenwich, CT: JAI Press Inc., pp. 317-339, 1984.

[172] R. Macwan, N. Patel, P. Prajapati, and J. Chavda, “A Survey on Various Techniques of Super Resolution Imaging,” International Journal of Computer Applications, Published by Foundation of Computer Science, New York, USA, Vol. 90, No. 1, pp. 19-22, Mar. 2014.

[173] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” IEEE Signal Processing Magazine, Vol. 20, pp. 21–36, 2003.

[174] P. Vandewalle, L. Sbaiz and M. Vetterli, “Signal Reconstruction from Multiple Unregistered Sets of Samples using Groebner Bases,” Proc. IEEE Conference on Acoustics, Speech and Signal Processing, Vol. 3, pp. 604-607, 2006.

[175] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” Int. J. Comput. Vision, Vol. 60, No. 2, pp. 91–110, 2004.

[176] H. Greenspan, C. Anderson, and S. Akber, “Image Enhancement by Nonlinear Extrapolation in Frequency Space,” IEEE Transactions on Image Processing, Vol. 9, No. 6, 2000.

[177] P. Vandewalle and J. Vandewalle, “Aliasing is Good for You: Joint Registration and Reconstruction for Super-Resolution,” Technical Report, 2006. http://lcav.epfl.ch/software/superresolution/index.html. [Online Accessed: May 2014]

[178] D. Keren, S. Peleg, and R. Brada, “Image Sequence Enhancement using Sub-pixel Displacement,” in Proceedings IEEE Conference on Computer Vision and Pattern Recognition, , pp. 742–746, June 1988.

[179] X. Xiao, J. Bahram, M. Manuel , and S. Adrian, “Advances in Three-Dimensional Integral Imaging: Sensing, Display, and Applications,” Applied Optics, Vol. 52, No. 4, pp. 546-260, 2013.

[180] A. Aggoun, "3D Holoscopic Imaging Technology for Real-time Volume Processing and Display," in High-Quality Visual Experience Springer, pp. 411-428, 2010.

[181] Z. Ni, D. Tian, S. Bhagavathy, J. Llach, and B. S. Manjunath, "Improving the quality of depth image based rendering for 3D Video systems," IEEE, Image Processing (ICIP), Cairo, pp. 513-516, 2009.

[182] D. Zarpalas, G. Kordelas and P. Daras, "Recognizing 3D Objects in Cluttered Scenes Using Projection Images," Image Processing (ICIP), 18th IEEE International Conference, pp. 673-676, 2011.

[183] S.-W. Min, B. Javidi, and B. Lee, "Enhanced Three-dimensional Integral Imaging System by use of Double Display Devices," Appl. Opt. Vol. 42, pp. 4186-4195, 2003.

[184] D.-H. Shin, M. Cho, K.-C. Park, and E.-S. Kim, "Computational Technique of Volumetric Object Reconstruction in Integral Imaging By Use Of Real and Virtual Image Fields," ETRI J, Vol. 27, pp. 208-712, 2005.

[185] D.-C. Hwang, K.-J. Lee, S.-C. Kim, and E.-S. Kim, "Extraction Of Location Coordinates of 3-D Objects from Computationally Reconstructed Integral Images Basing On A Blur Metric," Optical Society of America, Vol. 16, No. 6, pp. 3623-3635, 2008.

[186] A. Amar, Z. Dimitris, C. Paulo, and P. Peter, "3D VIVANT - Deliverable 4.1 - Accurate Depth Computation and Object Segmentation," 2011.

[187] N. Otsu, "A threshold Selection Method from Gray Level Histogram," IEEE Trans. SMC, Vol. 9, No. 1, pp. 62-69, 1979.

[188] http://www.mathworks.co.uk/products/demos/image/watershed/ipexwatershed.html#5. [Online Accessed: July 2014]

[189] R. C. Gonzalez, R. E. Woods, and S. L. Eddins "Morphological Reconstruction," http://www.mathworks.com/tagteam/64199-91822v00_eddins_final.pdf. [Online Accessed: July 2014]

[190] "Morphological Operations on Binary Images," Image Processing, Laboratory 7. http://users.utcluj.ro/~tmarita/IPL/IPLab/PI-L7e.pdf. [Online Accessed: July 2014].

[191] X. CHEN, S. LI, J. HU, and Y. LIANG, "A Survey on Otsu Image Segmentation Methods," Journal of Computational Information Systems, Vol. 10, No. 10, pp. 4287-4298, 2014.

[192] G. Li and Y. Wan, "Adaptive Seeded Region Growing for Image Segmentation Based on Edge Detection, Texture Extraction and Cloud Model," Information Computing and Applications Lecture Notes in Computer Science, Vol. 6377, pp. 285-292, 2010.

[193] N. Jamil, H. C. Soh, T. M. T. Sembok, Z. Abu Bakar "A Modified Edge-Based Region Growing Segmentation of Geometric Objects," Visual Informatics: Sustaining Research and Innovations Lecture Notes in Computer Science Vol. 7066, pp. 99-112, 2011.

[194] H.G., Kaganami and Z. Beiji, "Region-Based Segmentation versus Edge Detection," In: Fifth International Conference Intelligent on Information Hiding and Multimedia Signal Processing, IIH-MSP, pp. 1217-1221, 2009.

[195] T. Pavlidis and Y.-T. Liow, "Integrating Region Growing and Edge Detection," IEEE Trans. On Pattern Analysis and Machine Intelligence PAMI-12, pp. 225-233, 1990.

[196] B.S. Morse, B.S., "Brigham Young University," http://homepages.inf.ed.ac.uk /rbf/CVonline/LOCAL_COPIES/MORSE/threshold.pdf. [Online Accessed: June 2014].

[197] Z. Xiang, Z. Dazhi, T. Jinwen, and L. Jian, "A Hybrid Method for 3D Segmentation of MRI Brain Images," In: 6th International Conference on Signal Processing, Vol. 1, pp. 608-611, 2002.

[198] Yu, Y.-W., Wang, J.-H.: Image Segmentation Based on Region Growing and Edge Detection. In: IEEE International Conference on Systems, Man, and Cybernetics, Vol. 6, pp. 798-803, 1999.

[199] J. Fan, G. Zeng, M. Body and M.S. Hacid, "Seeded region growing: an extensive and comparative study," Elsevier, Pattern Recognition Letters, Vol. 26, pp. 1139-1156, 2005.

[200] http://cresspahl.blogspot.co.uk/2012/03/expanded-control-of-octaves- color map.html. [Online Accessed: June 2014].

[201] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and Practice of Background Maintenance" Proc. IEEE Int. Conf. Computer Vision, pp. 255-261, 1999.

[202] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," IEEE Transactions On Image Processing, Vol. 13, No. 11, November 2004.

[203] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge," http://pascallin.ecs.soton.ac.uk/challenges/VOC/. [Online Accessed: June 2014].

[204] T. Liu, H. Xu, W. Jin, Z. Liu, Y. Zhao, and W. Tian, "Medical Image Segmentation Based on a Hybrid Region-Based Active Contour Model," Computational and Mathematical Methods in Medicine, 2014. http://www.hindawi.com/journals/cmmm/2014/890725/. [Online Accessed: July 2014].

[205] http://www.mathworks.co.uk/matlabcentral/fileexchange/28149-snake-active-contour. [Online Accessed: July 2014].

[206] L. Goldmanna, F. De Simonea and T. Ebrahimia, "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic Video," Electronic Imaging (EI), 3D Image Processing (3DIP) and Applications, San Jose, USA, 2010.

[207] H. Benhabiles, G. Lavou´e, J.-P. Vandeborre, and M. Daoudi "A subjective Experiment for 3D-Mesh Segmentation Evaluation," Multimedia Signal Processing (MMSP), IEEE International Workshop, pp. 356-360, Oct. 2010.

[208] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet,"Evaluating Depth Perception of 3D Stereoscopic Videos," Signal Processing, IEEE Journal , Vol. 6, No. 6, pp. 710-720, 2012.

[209] ITU, "Methodology for The Subjective Assessment of the Quality of Television Pictures," in Recommendation BT 500-10, [Online Accessed: August 2014].

[210] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective Assessment Of H.264/Avc Video Sequences Transmitted Over A Noisy Channel," IEEE, Quality of Multimedia Experience, pp. 204-209, San Diego, CA, 2009.

[211] F. Šroubek, J. Flusser, and G. Cristobal, "Super-Resolution and Blind De-convolution for Rational Factors with an Application to Color Images," Computer Journal, Vol. 52, No. 1, pp. 142-152, 2009.

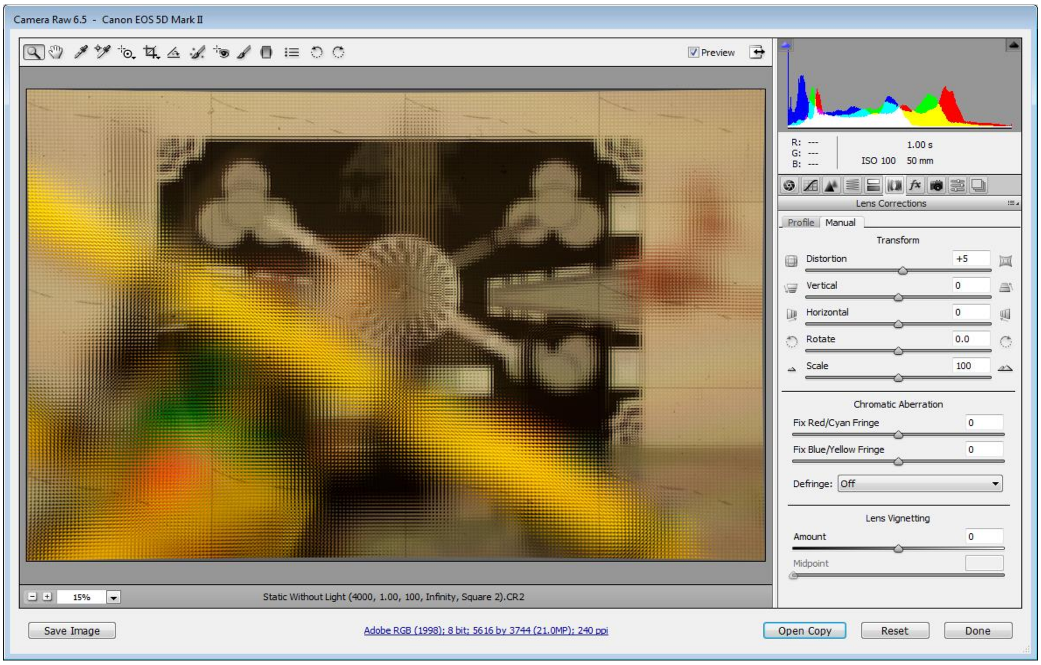
APPENDICES

Appendix A

Correct Barrier Distortion

The following steps are used to correct barrier distortion caused by the lens:

1. Right click on the RAW image data and open with Photoshop
2. Camera Raw window will open as can be seem in Figure A-1.
3. Click to the lens correction tab that appears on the right hand side in the camera Raw window.
4. Select manual that will give you the option of sitting the parameters manually.
5. In distortion selection box type +4 to fix the barrel distortion to its minimum.
6. After correcting the distortion click "open copy" button in the bottom right hand side in the window.
7. Then the image will open in Photoshop's window for cropping to remove the unnecessary borders to align all the micro lenses in vertically and horizontally directions as shown in Figure A-1.
8. After cropping save the image in .jpg or .png.



Camera Raw 6.5 - Canon EOS 5D Mark II

☑ Preview

R: ---
G: ---
B: ---

1.00 s
ISO 100 50 mm

Lens Corrections

Profile | Manual

Transform

Distortion +5
Vertical 0
Horizontal 0
Rotate 0.0
Scale 100

Chromatic Aberration

Fix Red/Cyan Fringe 0
Fix Blue/Yellow Fringe 0
Defringe: Off

Lens Vignetting

Amount 0
Midpoint

15%

Static Without Light (4000, 1.00, 100, Infinity, Square 2).CR2

Adobe RGB (1998); 8 bit; 5616 by 3744 (21.0MP); 240 ppi

Figure A-1: Illustration of Photoshop lens correction tool.

Appendix B

Correct Scaling Error Distortion

The block diagram of correction scaling error is presented in Figure B-1. The first step is generating the calibration grid from a white background of H3DI Figure B-2(a) and then this information is used to detect and eliminate dark borders introduced by the micro-lens array in the H3DI.

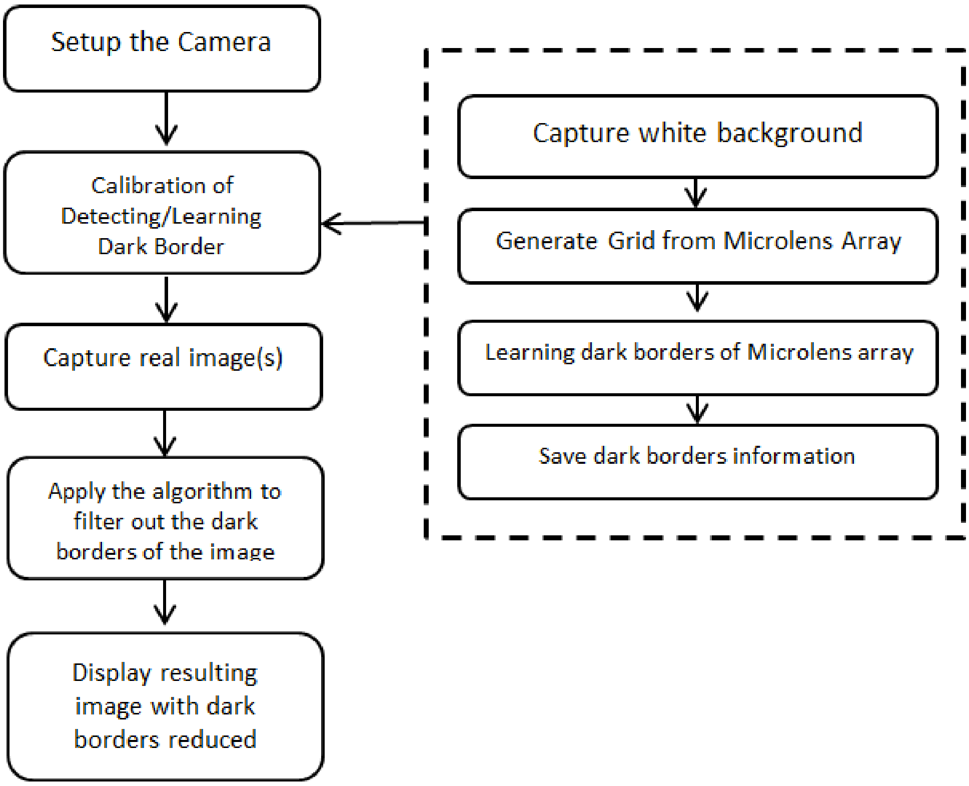
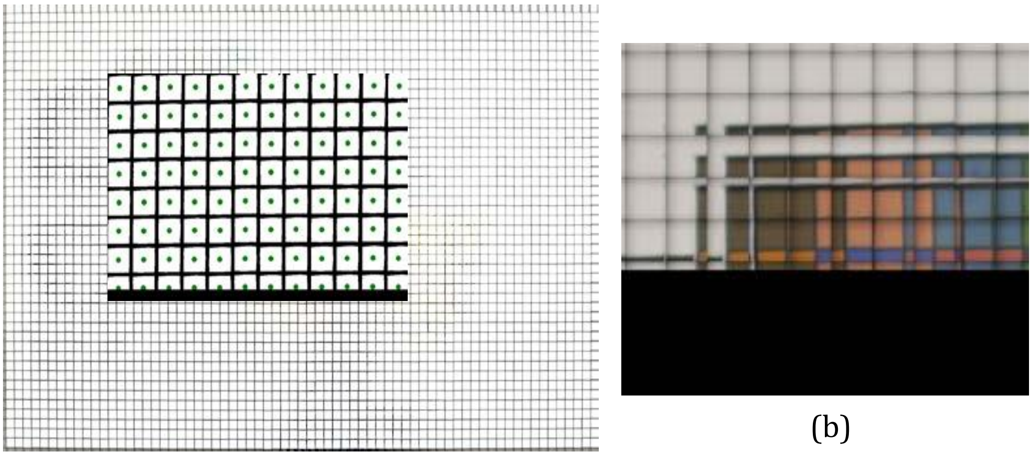

Setup the Camera
Calibration of Detecting/Learning Dark Border
Capture real image(s)
Apply the algorithm to filter out the dark borders of the image
Display resulting image with dark borders reduced
Capture white background
Generate Grid from Microlens Array
Learning dark borders of Microlens array
Save dark borders information

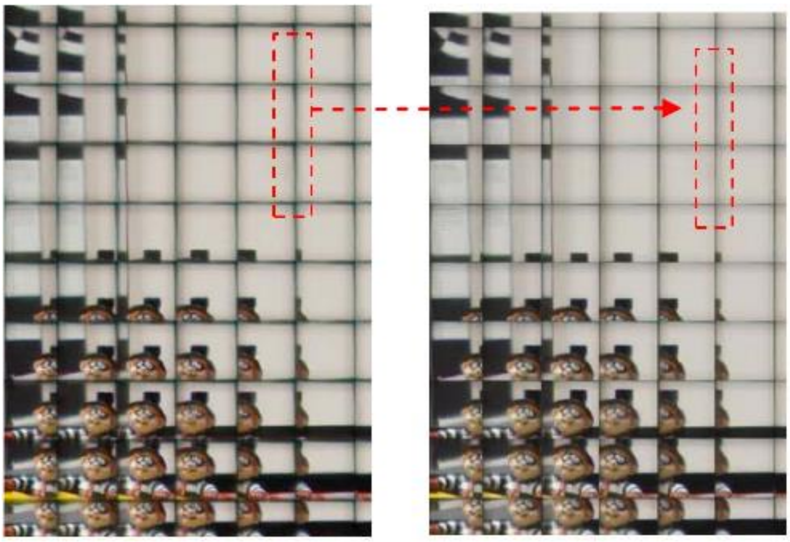
Figure B-1: Illustrates the block diagram of correction scaling error distortion.

The calibration grid enables one to accurately eliminate dark borders without effecting visual information. The eliminated dark borders of H3DI show in Figure B-2. The process does not remove the lens array border completely, but it refines enough the dark borders to reduce dark moiré. The reason of that, when removing the borders completely, this will also eliminate 3D visual details of EIs, and which should be avoided. The process is a self-tuning method, which is valid for all type of holoscopic 3D cameras. The resulting H3DI after correction process has refined micro-lens array borders shows in Figure B-3.



(a)

Figure B-2: Illustrates a) Generated calibrated micro-lens grid, b) Original image before correction.



(a)

(b)

Figure B- 3: Comparison of original OH3DI (a) with the resulting image, (b) after applying correction process.

Appendix C

Quality of Experience (QoE) Instructions

"In this experiment you will see different viewpoint images with different resolutions on your right hand side and left hand side. For each image shown you should judge its quality and choose a score on the continuous quality scale:"

Very low: "0-20" score if the image is very blurred with strong artefacts.
Low: "21-40" score if the image has many noticeable artefacts and very blurred.
Medium: "41-60" score if the image has some artefacts and medium level of blurring.
High: "61-80" score if the image is without any noticeable artefacts and there is a low level of blurring.
Very high: "81-100" score if the image is without any noticeable artefacts or blurring.

"Furthermore, in this experiment also you will see different 3D-interactive maps obtained from different 2D foreground object segmentation algorithms of the same viewed image on your right hand side screen, and on the left side you will see the depth map achieved from manual foreground object segmentation i.e. ground truth depth map, that is in front of you. For each display map, you should judge its quality and choose a score on the continuous quality scale:"

Bad: "0-20" score if the map has major boundary differences compared to the ground truth that destroy the object structure.
Poor: "21-40" score if the map has major boundary differences on some parts of the object compared to the ground truth.
Fair: "41-60" score if the map has minor boundary differences on some parts of the object compared to the ground truth.
Good: "61-80" score if the map has minor boundary differences on two parts of the object compared to the ground truth.
Excellent: "81-100" score if the map has minor boundary differences noticeable on one part of the object compared to the ground truth.

Thus, the schedule of the experiment is the following:

- Subject training phase (approx. 3 min)
- Break to allow time to answer questions from observers
- Test phase (approx. 30-35 min)

Assessment of results achieved from six holoscopic 3D images for each phase (four depth maps + one high resolution VPI).

Questionnaire

Test user sample characteristics

Test Number:	**Date:**

- **Age:**

- **Gender:**

- **What is your education status?**

• PhD Student	
• Master Student	
• Undergraduate	
• Others (Specify)	

- **What is your employment status?**

• Academic Staff	
• Researcher	
• Student	
• Others (Specify)	

- **Have you ever watched 3D movies?**

Yes □ No □

- **When was your first experience with 3D technology?**

• Years	
• Months	
• Weeks	

- **Have you got some experience in image processing (films, photos, etc)?**

Yes □ No □

Session 1: 3D Maps Quality Evaluation

Holoscopic Images	Bad	Poor	Fair	Good	Excellent	Methods
	0- 20	21- 40	41- 60	61- 80	81-100	
Horseman						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Tank						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Table						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Kidney						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Palm						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Plane						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Box-Tags						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
Airplane-man						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM
3DVIVANT						1st 3DIM
						2nd 3DIM
						3rd 3DIM
						4th 3DIM

Session 2: 2DVPIs Quality Evaluation

Holoscopic Images	Very Low	Low	Medium	High	Very High	Methods
	0- 20	21- 40	41- 60	61- 80	81-100	
Horseman						1st Image
						2nd Image
						3rd Image
						4th Image
Tank						1st Image
						2nd Image
						3rd Image
						4th Image
Table						1st Image
						2nd Image
						3rd Image
						4th Image
Kidney						1st Image
						2nd Image
						3rd Image
						4th Image
Palm						1st Image
						2nd Image
						3rd Image
						4th Image
Plane						1st Image
						2nd Image
						3rd Image
						4th Image
Box-Tags						1st Image
						2nd Image
						3rd Image
						4th Image
Airplane-man						1st Image
						2nd Image
						3rd Image
						4th Image
3DVIVANT						1st Image
						2nd Image
						3rd Image
						4th Image

Session 3: Users Feedback

- **What is your level of satisfaction with holoscopic 3D technology?**

• Extremely Satisfied	
• Satisfied	
• Dissatisfied	
• Not satisfied	

- **What is your general view of 3D technology?**

• Extremely Important	
• Important	
• Less Important	
• Not Important	

- **How you rate our quality of experience (QoE) assessment?**

• Extremely Important	
• Important	
• Less Important	
• Not Important	

- **How do you rate the simplicity/difficulty of the methodology of this assessment?**

• Very difficult	
• Difficult	
• Simple	
• Very simple	

Thanks for your time and participation. Hope you've enjoyed it.