CAMERA POSITIONING FOR 3D PANORAMIC IMAGE RENDERING

BY

ABDULKADIR IYYAKA AUDU

CAMERA POSITIONING FOR 3D PANORAMIC IMAGE RENDERING

A thesis submitted for the degree of

Doctor of Philosophy

by

Abdulkadir Iyyaka Audu

Supervised by: Prof. Abdul Hamid Sadka

Department of Electronic and Computer Engineering
College of Engineering, Design, and Physical Sciences
Brunel University
London

January 2015

To the loving memory of my mother and father I dedicate this work.

ABSTRACT

Virtual camera realisation and the proposition of trapezoidal camera architecture are the two broad contributions of this thesis. Firstly, multiple camera and their arrangement constitute a critical component which affect the integrity of visual content acquisition for multi-view video. Currently, linear, convergence, and divergence arrays are the prominent camera topologies adopted. However, the large number of cameras required and their synchronisation are two of prominent challenges usually encountered. The use of virtual cameras can significantly reduce the number of physical cameras used with respect to any of the known camera structures, hence adequately reducing some of the other implementation issues. This thesis explores to use image-based rendering with and without geometry in the implementations leading to the realisation of virtual cameras. The virtual camera implementation was carried out from the perspective of depth map (geometry) and use of multiple image samples (no geometry). Prior to the virtual camera realisation, the generation of depth map was investigated using region match measures widely known for solving image point correspondence problem. The constructed depth maps have been compare with the ones generated using the dynamic programming approach. In both the geometry and no geometry approaches, the virtual cameras lead to the rendering of views from a textured depth map, construction of 3D panoramic image of a scene by stitching multiple image samples and performing superposition on them, and computation of virtual scene from a stereo pair of panoramic images. The quality of these rendered images were assessed through the use of either objective or subjective analysis in Imatest software. Further more, metric reconstruction of a scene was performed by re-projection of the pixel points from multiple image samples with a single centre of projection. This was done using sparse bundle adjustment algorithm. The statistical summary obtained after the application of this algorithm provides a gauge for the efficiency of the optimisation step. The optimised data was then visualised in Meshlab software environment, hence providing the reconstructed scene. Secondly, with any of the well-established camera arrangements, all cameras are usually constrained to the same horizontal plane. Therefore, occlusion becomes an extremely challenging problem, and a robust camera set-up is required in order to resolve strongly the hidden part of any scene objects. To adequately meet the visibility condition for scene objects and given that occlusion of the same scene objects can occur, a multi-plane camera structure is highly desirable. Therefore, this thesis also explore trapezoidal camera structure for image acquisition. The approach here is to assess the feasibility and potential of several physical cameras of the same model being sparsely arranged on the edge of an efficient trapezoid graph. This is implemented both Matlab and Maya. The quality of the depth maps rendered in Matlab are better in Quality.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

Audu Abdulkadir, and Abdul Hamid Sadka. “Metric reconstruction of a scene from multiple images with single centre of projection.” Elsevier Visual Communication and Image Representation Journal. (under review).

Audu Abdulkadir, and Abdul Hamid Sadka. “Disparity map generation based on trapezoidal camera architecture for multi-view video.” International Journal of Multimedia and its Applications. (under review).

Audu Abdulkadir Iyyaka, and Abdul Hamid Sadka. "Generation of three-dimensional content from stereo-panoramic view." In ELMAR, 2013 55th International Symposium, pp. 101-106. IEEE, 2013.

Audu Abdulkadir Iyyaka, and Abdul Hamid Sadka. "Metric aspect of depth image-based rendering." In Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on, pp. 1-6. IEEE, 2013.

Audu Abdulkadir Iyyaka, and Abdul Hamid Sadka. "Computation of virtual environment from stereo-panoramic view." In 3DTV-Conference: The True Vision-Capture, Transmission and Dispaly of 3D Video (3DTV-CON), 2013, pp. 1-4. IEEE, 2013.

We have seen that vision is a complex phenomenon,
because it applies complex neural algorithm to bundle light rays so we see the world,
because it's principles are yet to be fully comprehended, and especially,
because it keeps humans and other advanced life forms on top of the food-chain.

ACKNOWLEDGMENTS

A straightforward expression of my gratitude is to say thank you. Perhaps this is strong enough perhaps not, that is the statement I can think of. The author would like to acknowledge the sublime mercy of Allah (subhaanawu wa ta alah). I never seriously felt unwell or was I short of sustenance during the course of my study. He made it possible for me to comprehend the solutions to diverse and complex research challenges. Prof. Abdul Hamid Sadka provided a refined level of supervision based on his breadth of academic experience . He was always welcoming whenever he was visited and I appreciate all the spirited discussion. Many thanks to everybody who already showed concern regarding my varied circumstances! At the Centre for Media Communication Research (CMCR), Brunel University, London, many thanks go to Mohib Hamdullah, Abdulfattah Obaid, Mohammad Rafiq, Muhammad Nawaz, Abdulkareem Ibrahim, Mohd Mohd Sazali, Salim Qatoob Al Amri, Saeed Alshahrani, Mussa Sheboniea, Salman Alharethi and Lawal Sagir.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
3DTV	Three-Dimentional Television
4D	Four-Dimensional
7D	Seven-Dimensional
BLAS	Basic Linear Algebra Subprograms
BA	Bundle Adjustment
CAPPA	Camera Array Pursuits for Plenoptic Acquisition
CCD	Charge Coupled Device
CTV	Conventional Television
CV	Covariance Variance
DIBR	Depth Image-Based Rendering
DOF	Depth of Field
DoG	Difference of Gaussians
DSIS	Double Stimulus Impairment Scale
EM	Electromagnetic
FOV	Field of View
FPGA	Field Programmable Graphic Array
FTV	Free-Viewpoint Television

GFSR	Generalised Feedback Shift Register
HD	High-Definition
HFM	Hierarchical Feature Matching
HPO	Horizontal Parallax Only
HVP	Human visual Perception
IBP	Iterative Back-Reprojection
IBR	Image-Based Rendering
IEEE	Institute of Electrical and Electronic Engineering
II	Integral Imaging
LAPACK	Linear Algebra PACKage
LDI	Layered-Depth Images
LMA	Levenberg-Marquardt Algorithm
LSF	Line Spread Function
MCP	Multiple Centres of Projection
MRF	Markov Random Field
MTF	Modulation Transfer Function
MV	Multi-View
MVS	Machine Vision System
MVV	Multi-view Video
NCC	Normalized Cross-Correlation
NDPE	Normally Distributed Pseudorandom Error
NDPPL	Normally Distributed Pseudorandom Pixels Locations

NSCP	Normalized Sum of Cross Product
OTF	Optical Transfer Function
PSF	Point Spread Function
RANSAC	Random Sampling Consensus
SAD	Sum of Absolute Differences
SBA	Sparse Bundle Adjustment
SD	Standard-Definition
SCOP	Single Centre of Projection
SCP	Sum of Cross Product
SIFT	Scale-Invariant Feature Transform
SFR	Spatial Frequency Response
SLAM	Simultaneous Localization and Mapping
SQA	Subjective Quality Assessment
SQF	Subjective Quality Factor
SSD	Sum of Square Differences
SURF	Speeded Up Robust Features
TCA	Trapezoidal Camera Architecture
TGFSR	Twisted Generalised Feedback Shift Register
TOMBO	Thin Observation Module by Bound Optics

Part I

INTRODUCTION

1 INTRODUCTION

1.1 INTRODUCTION

This chapter provides an introduction to visual content acquisition in relation to multi-view video using multiple cameras. In order to reduce the number of physical cameras used, both the field of image-based rendering and camera positioning which are the major objects of the work presented in this thesis are highlighted. Starting with what is known about human vision system, consideration of the motivation, aim and main objects of this research are discussed. Furthermore, the major research contributions and methodology used are described. Finally, the structure of this thesis is outlined.

1.2 BACKGROUND OF THE THESIS

Humans and other advanced life forms effectively interact with the environment through the development of sophisticated sensory methods. Vision is one of these sensory methods. Vision is widely understood to provide sufficient information in order for humans to infer the optimal response under a variety of situations [1]. An identification of first degree is that the Human visual Perception (HVP) under the control of complex neural algorithm provides visual information in 3D form at high resolution and colour [1].

It is revealed that overcoming distance within a certain range both in space and time has remained an important characteristic of HVP. Precisely speaking, there is no time lag in seeing a point close to the point of observation and a point which is further. This feature according to [2] has been significantly demonstrated and successfully used in some applications such as Conventional Television (CTV) which is known to be an outstanding visual media in recent time. It allows a distant world to be seen in real time. However, it has to be mentioned that CTV is limited by a lack of suitable 3D information.

Depth perception is another subtle dimension in the understanding of the integrity of HVP. Humans have two eyes with an interocular distance of about 7cm and as a result each eye sees a slightly different view of the scene. In this way, the HVP exhibits retinal disparity. The reciprocal of retinal disparity is known as depth and gives an estimate of how far different points in the scene are with respect to the point of observation. The investigation of the roles of binocular neurons in different perceptual tasks in the view of [3] has advanced an understanding of the stages within the visual cortex that creates an internal representation of the world using different depth cues leading to binocular depth perception. Therefore, most natural phenomena are expressed as a function of these depth cues. Such depth cues include: object relative size, visibility, speed, and occlusion. Of course, we can also very easily answer questions such as: Is that lion looking at me at a far distance?, Is that stone really flying towards my head?, and Can I jump over this ditch? [4].

Further insight into HVP has been made possible by the established developments in the understanding of horizontal parallax. For a given 3D object scene, as briefly mentioned earlier, a slightly or considerably different view is observed whenever the viewpoint with respect to the scene is changed. This characteristic greatly improves the viewing experience of the observer. In robotics navigation, object tracking, video surveillance [5], it is increasingly common to replicate as nearly as possible some of the characteristic of HVP and most especially in multimedia to provide an extended view of the environment in advanced visual media devices of which Three-Dimentional Television (3DTV), Free-Viewpoint Television (FTV) are important and have an extra ordinary standing. They are expected to revolutionize the culture and entertainment industry and consumer electronics industry. The key idea behind the generation of 3D content is to be able to provide the viewer with an illusion of depth as seen in the real world [6]. Of course, some important functional blocks are needed to be able to achieve this with high performance and reliability.

At one extreme end, it is essential that the television is well equipped with auto-stereoscopic display to provide the observer with 3D images without requiring special glasses [7–10]. The design of auto-stereoscopic displays has enjoyed extensive research attention driven by the expectation of 3D video to become the mode of visual content communication on the internet [11]. In 3D video, the full 3D shape, motion, and surface texture are documented. Among various auto-stereoscopic display techniques [12], Multi-View (MV) display has a top ranking because of the more effective and easy of implementation. With MV display, the future offers the possibility of high quality 3D images [13, 14] at several pre-defined viewpoints. This is achieved through a special ray guiding optics such as parallax barrier or lenticular lens sheet [15].

Many other high performance display technologies are known to exist. Integral Imaging (II) is a whole new class of auto-stereoscopic display strategy. II display is understood to constructs 3D images based on the reconstruction of spatio-angular ray distribution using an array of lenses. Comparatively speaking, the design of MV displays are to provide different viewpoints with different views. However, in II display, the observer space is manipulated such that the ray distribution is reconstructed. In this way, II display features a substantial improvement of natural and continuous motion parallax characteristics. Also, the commercialisation of II display has received enormous attention due to the specific advantages mentioned earlier. The major drawback of insufficient panel resolution is addressed in these display methods at the expense of vertical parallax and a reduction in the number of rays per each display cell to the level comparable to that of MV displays. The Horizontal Parallax Only (HPO) II display is now known to be another auto-stereoscopic display technique that has been widely embraced. However, with recent advances in theory and micro-lens manufacturing, II can effectively be adapted to an auto-stereoscopic display using using bidirectional 3D II [16–18]. The other benefit of this design is that each view has a better aspect ratio; instead of splitting the display horizontally into many views, both horizontal and vertical directions are split. The slanted lenticular arrangement requires sub-pixel mapping to image all pixels along a slanted line in the same direction.

On the other extreme, the strategy of content acquisition which is an important aspect of this research work, must be effective enough to provide an accurate optical and geometrical registration of scene objects. This will allow a wide range of applications. The performance of auto-stereoscopic displays is considerably affected by image distortion. The acquisition of image contents with correct geometry is a fundamental requirement in order to avoid any distortions of 3D images at the display. Investigations based on experimental studies have indicated that the contents capturing condition using multiple cameras or computer graphic techniques can deviate from the ideal one, resulting in distortion of the displayed 3D images. This problem is more severe when HPO II display is considered. Another important observation concerns the image contents orthographic requirement of II display. The usual perspective image contents captured using cameras cannot be used in principle. Hence there is scope for the scene capture for Multi-view Video (MVV) to be improved without much computational cost in data processing.

Visual content can be created through the use of cameras or computer graphic procedures. Of recent, there is evidence of the continuous evolution of devices to capture images in its desire size and optical components [19–21]. Also computational photography is now highly valued in the capturing, processing, and manipulation of images that improve the capabilities of photography using computers. The technological frontier of image acquisition is currently being set by light field cameras. But an improved understanding about light field cameras indicates that the low-resolution images are generated due to their spatio-angular light field sampling. Small output image is usually produced in comparison with sensor size since the angular information component is known to degrade spatial resolution. Though super-resolution techniques have been advanced to address the spatial resolution issue, such approaches are not suitable for real-time application. Also, [22] has observed that the development of algorithms that process light fields is lagging behind. For instance, synthetic refocussing alone may be insufficient to establish light-field photography in the mass market.

The use of dedicated hardware that increases the computational processing power is currently understood to be the solution for real-time processing. In recent years, [19], Field Programmable Graphic Array (FPGA) have been used to perform general purpose computations in sensor development for telecommunications, networking, or consumer and industrial applications with a significant speed-up. The low cost of the FPGA implementation and its low-consumption of energy makes this solution attractive for an implementation embedded in plenoptic cameras.

However, the systematic combination of several physical cameras in a particular topology continues to be recognised to offer an opportunity to improve the quality and realiability of the acquired image [23]. The fundamental decision to use dense or sparse camera arrangements is usually critical in many applications. Dense camera arrangement provides considerable overlap and hence it is highly favoured despite the implementation cost. To reduce cost and some other associated problems with dense camera arrangement, the use of virtual cameras has been promoted. This involves the use of image processing to interpolate the pixel information acquired by certain number of physical cameras.

Therefore, content acquisition using multiple real and virtual cameras is highly challenging and has become an enabling technology. The ability to reconstruct scene optical properties from a series of image measurements is extremely valuable in a range of applications as it potentially allows visual technology with superior viewing selectivity. This is expected to be the basis of the research work presented in this thesis.

1.3 MOTIVATION

A very important factor for successful 3D image acquisition is camera positioning. Parallel, convergence, and divergence array are the three well-established conventional camera architectures that have been described in literature and routinely used in film industry for characterisation of 3D video production. They impact significantly on the integrity of post production video content. Any suspected problems are usually addressed during post-production process. As an example, the mutual occlusion of multiple objects in motion can degrade the quality of 3D video produced [24]. As a result, the current 3D capturing method based on multiple cameras arranged in one of the conventional topologies assumes a single object in motion in a well-designed studio in conjunction with a consideration of the dynamic of lighting environments system. Also, a recent study by [25] has shown that the presence of discontinuities in images is a concern especially in emerging multi-video applications where the service integrity of movies is a fundamental requirement to avoid viewing discomfort and eye fatigue.

However, camera positioning is difficult to realise in cases where the calibration requirements both geometric and photometric are not satisfactorily met with a reasonable degree of accuracy. Also for the purpose of 3D reconstruction, it is highly required that the whole surface area of any scene object be visible in at least two cameras [24]. Based on the known simple mathematical model and analysis of the existing camera architectures, it has been shown that this requirement is a function of camera density [26]. Excellent simulation results documented in literature indicate that dense camera arrangement as opposed to sparse arrangement would have been the most suitable solution for scene object visibility. This target is challenging due to the complexity of the profiles encountered in practice. Dense camera arrangement is known to be closely accompanied with synchronisation problem. In practice, MV cameras may not be simultaneously triggered. So it would be very useful to apply image-based rendering techniques to synthesize virtual cameras in-between physical cameras. Significantly, this idea is an active research concept. It relies on the both the geometric and optical data acquired by few physical cameras. Another challenge of significant proportion is that the interocular relationship between any two physical cameras must be known.

A prominent feature of the existing camera architectures is that they are planar. All cameras are usually constrained to the same horizontal plane. A camera architecture in which some of the cameras are placed in different horizontal planes might have the potential to significantly address some problems such as occlusion which is known to pose serious issues in content creation.

1.4 AIM AND OBJECTIVES

In multimedia, a network of cameras is often used in the creation of 3D content. In spite of the known performance issues, most of the cameras fabricated using the current state-of-the-art technology have the unique potential to provide assurance of the viewing integrity of the completed 3D content. However, any completed 3D content may still host certain degree of issues related to viewing discomfort and eye fatigue, and occlusion. Hence the structure of the network of cameras is crucial. Currently, 3D content creation is carried out using either parallel, convergence or divergence camera arrays with dense or sparse characterisation. With any of these arrangements, all cameras are constrained to the same horizontal plane. Therefore, occlusion becomes an extremely challenging problem and robust camera set-up is required in order to resolve strongly the hidden part of any scene objects. To adequately meet visibility condition for scene objects and given that occlusion of the same scene objects can occur, a multi-plane camera structure is highly desirable. Trapezoidal Camera Architecture (TCA) is to be explored for image acquisition. The present work is an attempt to enhance the understanding of trapezoidal camera architecture from the fundamentals of basic trapezoid and to determine the controlling parameters of this phenomenon. The overall analysis starts with a study of the image-based rendering which needs no geometric model and eventually realize a virtual camera. Using the definitions from previously published literature materials and well-established understanding of image processing, this thesis aims to realize a virtual using the concept of image-based rendering. This aim will lead to the following objectives:

- To study the effect of region match measures on quality of rendered image.
- To generate a 3D content from a pair of panoramic images.

- To perform metric scene reconstruction based on sparse bundle adjustment algorithm.
- To study the possibility of trapezoidal camera structure.

Matlab, Meshlab, and C programming languages among others will be employed in the implementation of the aforementioned aim.

1.5 CONTRIBUTION TO KNOWLEDGE

The following are the knowledge contribution of the research work presented in this thesis:

A. The effect of pixel-based matching cost namely absolute differences, squared differences, normalized cross correlation has be investigated in chapter four, section 4.1, in a work titled "Metric Aspect of Image-based Rendering" and presented at the conference 2013 1st International Conference on Communications, Signal Processing, and Their Applications (ICCSPA), American University of Sharjah, UAE, 12-14 February 2013.

B. 3D content generation from stereoscopic panorama is another important contribution of this thesis. Depth perception is achieved by super imposing two slightly different panoramic views of the same scene. In the multiple view acquisition stage, two Nikkon D7000 cameras with a stereoscopic distance of 150mm were mounted on the same tripod. Depth perception is observed using a pair of anaglyph glasses. The work presented in chapter four, section 4.3 is based on this concept. This has been published in the proceedings of the 55th International Symposium, ELMAR-2013, which took place on 25-27 September 2013, Zadar, Croatia.

C. Also, panoramic depth image-based rendering has been carried out. It aims at constructing a new view by texturing the depth image obtained from two stereo panoramic views. The implementation of this can be found in section 4.5 of chapter four. This work featured in the 2013 3DTV Conference: Vision Beyond Depth, Aberdeen, UK, 7-8TH October 2013.

d. Camera position and orientation in a MV setting has been investigated based on a randomly generated synthetic scene. classic bundle adjustment algorithm has been employed in the analysis. The performance is a fair one since the scenes considered are not realistic. However, the investigation has provided an insight into camera position estimation. The detail of this can be found in chapter five, section 5.1. Also, experiment on reconstruction of 3D scene from multiple Two-Dimensional (2D) image samples of the same scene has been performed in section 5.3. The image samples used have Single Centre of Projection (SCOP). The generic Sparse Bundle Adjustment (SBA) algorithm used for this type of work does not make available the optimised 3D coordinate points. In this work, this trend has been modified. It is now possible to access the optimised values and use them in a further stage depending on the need. This is significant since, the use of this algorithm has been avoided in the past because what has been optimised could not be accessed. Only text information stating the number of points optimised, the number of iterations performed, and complexity in terms of how long it took for the process to be completed.

e. The formulation of trapezoidal camera architecture TCA for MVV is covered in chapter six. In conventional camera topologies namely: linear, convergence, and divergence, all the cameras are confined to the same horizontal plane. However, in TCA, participating cameras can have different vertical coordinates. This concept provides the opportunity for all the points in the scene especially the occluded ones to be observed from several viewpoints on the edge of the trapezoid. Furthermore, the idea has also been simulated and evaluated in Maya in order to establish its viability. The ideas from the concept of TCA have been used to construct high quality dynamic depth map for graphic images. Both the mathematical description of TCA and the depth map generated as a result have been rigorous and impressive.

1.6 RESEARCH METHODOLOGY

The object of the research presented in this thesis is achieved based on the following research methods:

- Image-Based Rendering (IBR) with and without geometry has been chosen as a research method since it is easier to solve correspondence problem with multiple image samples of the same scene. With multiple image samples it is also possible to perform self-calibration where both the intrinsic and extrinsic parameters are extracted from the images. The rendered or reconstructed scene is more complete when multiple image samples of the same scene are used in the processing chain.

- For the metric reconstruction from multiple image samples, SBA method is favoured since it is not computational expensive to optimise the set of re-projected 3D points which constitute an object. This is as a result of the fact that there is no interaction among parameters for different 3D points and cameras. Hence, the underlying normal equations exhibiting a sparse block structure.

- The properties of a trapezoid graph have been explored in the proposed trapezoidal camera architecture. These properties easily allow any two cameras located on the edge of the trapezoid to be related. This leads to the computation of the coordinates of the cameras. Consequently, a virtual viewpoint on the edge of the trapezoid can be calculated.

- This research started by establishing the state-of-the-art in image-based rendering and camera pose estimation. A large volume of literature material with broad relevance such as books, conference and journal papers were thoroughly revised. A deep-rooted understanding of the work carried out in the relevant papers studied, was based on knowing the research gap being addressed and the method used in its realization. Journal papers with high impact factor were mainly considered. These include journal papers from Institute of Electrical and Electronic Engineering (IEEE), Elsevier, and Springer journals.

- Books were also consulted. Such books have the principal objectives of providing an introduction to basic concepts and methodologies for digital image processing, and develop a foundation that can be used as the basis for further study and research in the field of image processing. Key chapters on feature detection, extraction, and matching in such books were studied.

- Preliminary investigation on the application software best suitable for the set aim and objectives was made.

- Various training sessions organised by Centre for Media Communication Research (CMCR), College of Engineering, Design and Physical Sciences, Graduate School, and National Instruments were attended.

- The theoretical concepts used during the course of this research were implemented using MATLAB and C programming languages. Meshlab, photoshop, 3ds max, and Imatest were also used intermittently to either modify, visualise, or make quality assessment of some of the input images or simulation results.

- Constant consultation with my supervisor and fellow researchers to discuss research areas and issues which needed clarity.

- In order to gauge the obtained results, known and reliable performance parameters were used for different segments of the research contribution. These performance parameters give an indication of the strength and weakness of the proposed method over known standard.

- Research results were presented in local, national and international conferences. The conferences provided opportunity for the research results to be assessed by other experts in the field of image processing and other related fields. Some of the results were improved upon and published in high impact factor journal.

1.7 THESIS STRUCTURE

The work presented in this thesis is arranged as shown in Figure 1.1. It consists of Four parts namely: Part I, Part II, Part III, and Part IV.

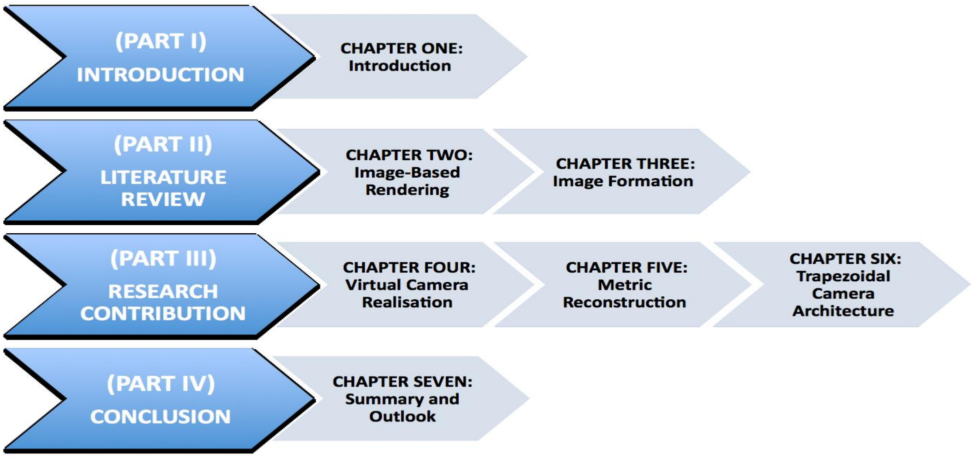

(PART I)
INTRODUCTION
CHAPTER ONE: Introduction
(PART II)
LITERATURE REVIEW
CHAPTER TWO: Image-Based Rendering
CHAPTER THREE: Image Formation
(PART III)
RESEARCH CONTRIBUTION
CHAPTER FOUR: Virtual Camera Realisation
CHAPTER FIVE: Metric Reconstruction
CHAPTER SIX: Trapezoidal Camera Architecture
(PART IV)
CONCLUSION
CHAPTER SEVEN: Summary and Outlook

Figure 1.1: Outline of thesis structure.

In Part I, Chapter One is presented.

- Chapter One contains the "Introduction" of the thesis. It provides a straightforward introduction about this thesis. Key components such as background of study, motivation, aim and objectives, research methodology, research output are all clearly spelled out.

Part II covers "Literature Review" and consists of Chapter Two and Chapter Three

- Chapter Two is titled "Image-based Rendering" A review of the fundamentals of known camera architecture and image-based rendering are highlighted. This ranges from content acquisition principle and techniques, to content processing and visualization.
- Chapter Three is titled Image formation. Image formation process is presented from the perspective of wave optics. This is highly necessary since this

research is on content acquisition and its processing. Also, this modelling leads to an objective quality assessment parameter. This parameter is an important camera or image quality parameter known as Modulation Transfer Function (MTF).

The original contributions of this research are presented in Part III. Chapters: FOUR, FIVE and SIX are contained there.

- Chapter Four is titled "Virtual Camera Realisation". It contains three sections with the following subtitles:

1. Metric Aspect of Image-based Rendering.
2. Generation of Three-Dimensional Panoramic Image.
3. Computation of Virtual Environment from Stereo-Panoramic Images

- Chapter Five is titled "Metric Reconstruction". This chapter covers reconstruction of scenes from several multiple images of the same scene with SCOP. SBA algorithm is the method of focus.

- The title of Chapter Six is "Trapezoidal Camera Architecture". Camera structure aimed at acquisition of 3D content is discussed. Part IV is where the summary and conclusion of this research is drawn.

1.8 CHAPTER SUMMARY

In this chapter, the judgement of the significance of this thesis can be found. It follows from the fact that a dense camera structure at the acquisition end of a 3D pipeline can enhance the reality and naturalness of 3D image at the consumer end. Two major challenges in close proximity with this idea are the use of large number of physical cameras and their synchronisation. However, it is important that the number of physical cameras used is minima. This means that virtual cameras can form an integral part of dense camera arrangement aimed at visual content acquisition for multi-view video. Virtual cameras can be reliably constructed through the method of IBR. Hence, a virtual camera is used in this thesis to render a scene with respect to a viewpoint by depth map texturing, to generate 3D content from stereo panoramic image. Computation of panoramic virtual environment using depth map method is also implemented. Starting with multiple image samples, metric reconstruction is realised based on SBA algorithm. Finally, a trapezoidal camera architecture is proposed. This is a multi-plane configuration in which all cameras are not constrained to the same horizontal plane.

Part II

LITERATURE REVIEW

2

IMAGE-BASED RENDERING

2.1 INTRODUCTION

This chapter establishes the state-of-the-art in image-based rendering and camera structure for three-dimensional content acquisition. A tailored review is given with respect to each intended research contribution. This is pursued in order to bring to lime-light an understanding of the rendering capabilities of virtual camera, as well as to aid the image processing procedures used for its realisation. It also provides for critical analysis which leads to the definitions of the research gaps. It starts from historic perspective with an overview of stereopsis.

2.2 BACKGROUND

There has been a successful exploitation of the basic principle of stereopsis [27] first demonstrated by Sir Charles Wheatstone in 1838 [28]. Stereopsis is the perception of depth and 3D structure obtained on the basis of visual information derived from two eyes by individuals with normally developed binocular vision [29]. Stereopsis is understood to depend on the disparity between views that a 3D object affords to each eye. This binocular disparity can be perceived as a horizontal shift in the visual field of one eye compared to the other when each eye is closed in alternation. The two images are fused in the cerebral cortex and experienced as a single 3D representation under normal circumstances [30, 31].

After the laboratory curiosity to record light came to fruition, the process to replicate stereopsis in visual media became an active area of research. This is evident in the work of Alfred Molteni [32] shown in Figure 2.1. Further devel-

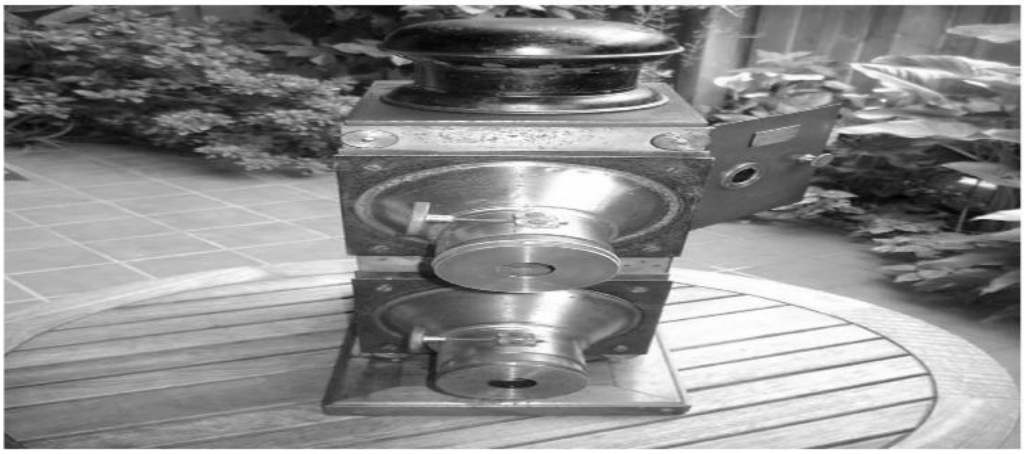
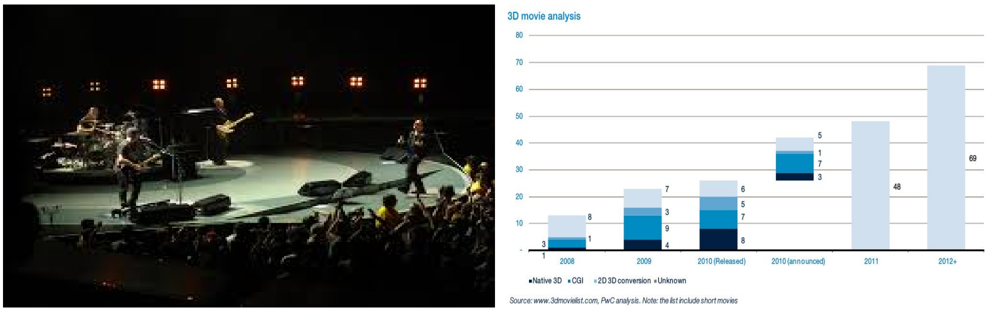


Figure 2.1: Alfred Molteni's biunnial magic lantern with two vertically stacked projec tion lenses, used to project anaglyph images (1890) (photo by Ray Zone; Erkki Huhtamo Collection). Taken from [32].

opment was facilitated by the considerable improvement in the computational capability of microcomputers which is attributable to the advancement in semiconductor technology. As a consequence, this has brought about the convergence of two major fields of research: Computer vision and computer graphics. In fact, it is argued that the continuing developments in camera fabrication and image processing will further augment the already high applicability and versatility of stereopsis techniques. Therefore, it can be expected that optical techniques will continue to gain importance in many fields of application. Understandably, several other application areas with extensive possibilities have emerged which try to explore further the successes of earlier fields of research. Some of these areas of application have been driven by the need to optimise information acquisition and analysis.

The degree of accuracy expected by a wide range of emerging applications which depend on digital images as input has resuscitated the challenge and significance of content acquisition. In particular, the acquisition of 3D content, Figure 2.2 (a), based on the statistics shown in Figure 2.2 (b), still lags behind 3D display technology. In multimedia, 3DTV and FTV illustrated in Figure 2.3 are devices which require accurate digital images in order to provide the necessary viewing experience [33–41]. 3DTV provides depth illusion while FTV allows for immersive



Source: www.3dmovielist.com, PwC analysis. Note: the list include short movies

(a)

(b)

Figure 2.2: (a) Live 3D production taken from http://www.google.co.uk/search?q=3d+live+production, (b) 3D movie releases 2008 to 2012.

experience through the choice of viewpoint. These systems fundamentally try to replicate the HVP through the use of stereopsis and motion parallax respectively. Currently, the existence of prominent 3DTV channels has been confirmed. Few examples are: The satellite BS11 (airs 3D content four times a day); South Korea: Sky 3D; United Kingdom: BSkyB Channel 3D, Virgin Media VoD; United States: New York Network Cablevision, Discovery 3D, ESPN 3D, DirecTV VoD service; Australia: Fox Sports; Russia: PlatformHigh-Definition (HD) in partnership with Samsung; Brazil: RedeTV; France: 3D broadcasts by Orange, Numericable, and Canal+; Spain: Canal+ [42]. A live 3D production is shown in Figure 2.2 (a).

A 3DTV provides depth illusion by beaming out of its screen several views of a scene which are slightly different and the eyes of the observer pick up any two consecutive views. From the perspective of HVP, a 3DTV can only provide MVs from its screen if several images have been previously acquired with multi-cameras strategically positioned around the scene [43–47]. Camera configuration around a scene may go well beyond parallel camera array such as in Figure 2.4 (a). Therefore, even though dense camera arrangement using other established topologies could reduce the re-projection error on the 3DTV and FTV screens, the cost estimate, geometrical calibration, colour balance between the individual cameras, mechanical limitation, and temporal synchronisation issues may not make it viable in some critical applications. Hence it does clearly indicate that either the sophistication or density of acquisition cameras will have to be en-

3D
PENINSULA
周回型スタジアム自由視点映像
www.diginfo.tv
Until now, the viewing angle for live soccer has only been changeable by switching between camera positions.
REPUBLICAN PRESIDENTIAL DEBATE
#cnndebate
LIVE CNN

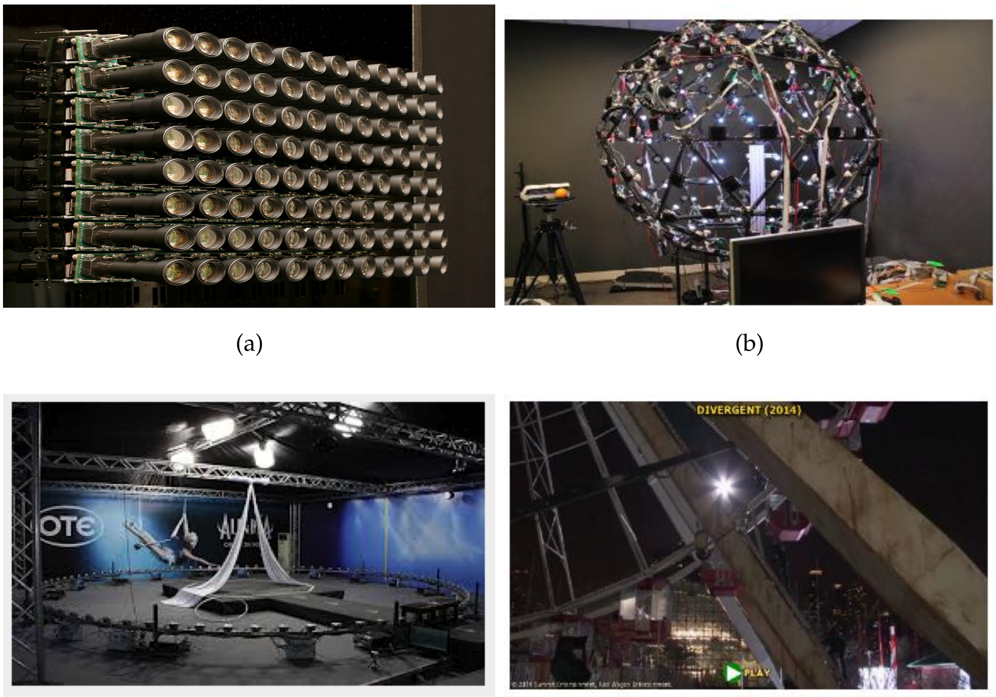
(a) (b)

(c) (d)

Figure 2.3: Illustration of 3DTV (a), and FTV (b), (c), (d). Taken from http://www.google.co.uk/search?q=3dtv.

hanced in conjunction with a careful investigation of all cost issues. An alternative technique is to have the cameras sparsely arranged in certain topology depending on the need [48–50].

In order to make the sparse camera arrangement more useful, the use of virtual camera has evolved into a central challenge in multimedia. Virtual cameras could be created in between the physical cameras. The concept of virtual camera is primarily directed towards a photo-realistic synthesis of images using reference images acquired by the few available physical cameras. The immense potential of virtual cameras extends beyond the sphere of influence of multimedia. It is also commonly applied in augmented reality, robotics and navigation. It can be realised through the sampling and interpolation of the pixels contained in the images acquired by the physical cameras. This technique is referred to as IBR [38, 51–58]. Therefore, IBR is a sampling and interpolation problem. In IBR concept, the collection of light rays emanating from the surface of an object constitute the scene [59]. The MV image set, therefore, samples this representation of



(a)

(b)

(c)

(d)

Figure 2.4: Typical camera arrays. Taken from http://www.google.co.uk/search?q=camera+array

the scene with each image recording the intensity of a set of light rays travelling from the scene to the camera. The implementation of IBR involves the reduction in the dimensions of plenoptic function [60].

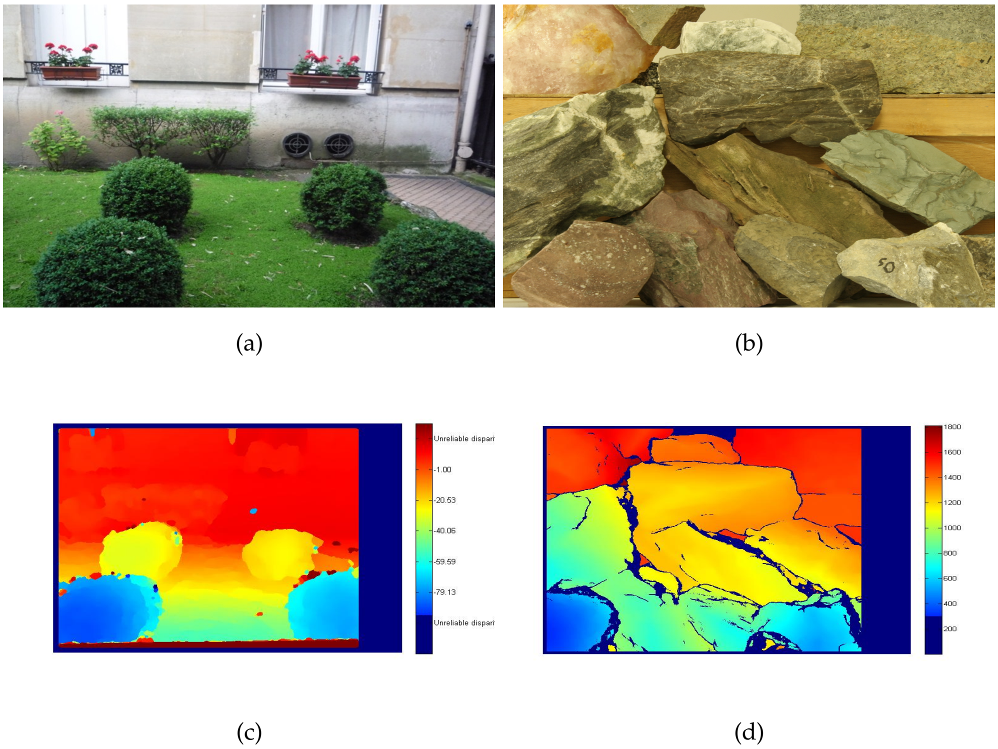
2.3 CATEGORISATION OF IMAGE-BASED RENDERING

Of course, pixel indexing scheme has been advanced by [61], for categorising the implementation methods of IBR, the criteria based on the amount of geometry required is widely accepted in the computer vision research community. Therefore, the discussion in this thesis will be based on the geometry criteria.

2.3.1 *Geometry Specific Rendering*

Various methods of IBR, going well beyond rendering with explicit accurate geometry and not much number of images [62–71]. In this method, a model is developed from multiple image samples of the scene and the corresponding depth map. The depth map shown in Figure 2.5 contains the depth associated with every pixel of the reference images. Hence, by re-projecting the set of 3D points, the complete image can be rendered with respect to a virtual viewpoint. However, construction of depth map is the major challenge associated with 3D video which uses images from dense multiple viewpoints to provide the sense of immersion. This is because there exist intricate geometric relations between multiple views of a 3D scene [72]. These relations are related to the camera motion and calibration as well as to the scene structure. As noted by [73], future development in the direct capture of depth-enhanced 3D video will have to be improved upon. Currently, depth sensor enhanced camera setups are rarely used, because of the limited spatial resolution and depth range of available sensors.

Thus, the acquisition of high-quality depth maps is paramount in 3D video and related applications. When the camera intrinsic parameters are known, motion algorithms have the potential to use distinct feature points to generate surface models with sparse characteristics. However, the reconstruction of a scene which is geometrically correct characterised with visually pleasing surface models requires more than just knowing the camera intrinsic parameters [72]. Re-



(a)

(b)

(c)

(d)

Figure 2.5: Scenes and their corresponding depth maps.

construction is realised by a dense disparity matching process that computes correspondences from the grey level images directly by exploiting additional geometrical constraints. Therefore, dense surface estimation is achieved in a few number of steps. First image pairs are rectified to the standard stereo configuration. Image rectification explored epipolar constraint to convert the correspondence search problem from 2D to a search along a straight line (i.e One-Dimensional (1D)). This is because corresponding feature points in stereo image now have the same vertical coordinate and different horizontal coordinates. Second stereo matching algorithm is used to generate disparity maps as will be mentioned in the next paragraph. In the third step, the available pairs of view are made available for integration based on MV approach.

Correspondence estimation in stereo vision has two broad levels depending on the search constraint imposed [74]. Local constraint is usually exercised on few pixels around the pixel of interest. Feature matching, gradient-based methods, and block matching techniques use local constraints and are known to be efficient in the face of sensitivity to occlusion regions, texture-less regions or regions with uniform texture [75–84]. Feature matching methods have been widely explored. Two main ideas are responsible. Dense depth maps are critical for a variety of applications. Also, the availability of promising and robust regression algorithms. This is on the merit of the efficiency of feature matching methods to handle depth discontinuities. This is because a region of support near a discontinuity contains points from more than one depth. It is also known to adequately deal with regions of uniform texture in images. In both of these challenges which are known to be the drawbacks of gradient and block match methods, feature-based methods robustly act at the expense of the density of points for which depth may be estimated, by reducing the regions of support to specific reliable features in the images [74].

Driven by the need for high-quality depth maps in certain critical applications, segmentation matching and Hierarchical Feature Matching (HFM) now constitute a new dimension of advanced feature-based method for stereo correspondence [85–87]. HFM exploit lines, vertices, edges, and edge-rings. The strength of HFM is that it favours the use of coarse, reliable features in order to provide support for matching finer, less reliable features, and it minimises the computational complexity of matching by reducing the search space for finer levels of features. In segmentation matching, the method involves iterative partitioning of stereo images. This is then followed by the calculation of affine transformation matrix of six parameters which is used to model the relationship between the small regions. Based on the affine matrix, matching of the segmented regions of stereo images is performed.

For global constraint, the search area is the entire image. It comprises of techniques like belief propagation, graph cuts, non-linear diffusion, intrinsic curves, correspondence-less methods, and dynamic programming [74, 88–98].

Depth sensors based on the time-of-flight technique [99] have proved to be promising in its ability to acquire the depth map of the scene and is now well advanced. One main factor which has slightly obscured its more widespread use for depth map acquisition in real-time is its low resolution output. Depth map up-sampling methods such as filter-based, Markov Random Field (MRF)-based, structure guided fusion techniques have been proposed [100–111], to resolve this challenge and other related issues. Typically an MRF-based stereo vision algorithm employs a likelihood function that reflects the local similarity of two regions and a potential function that models the continuity constraint. As a consequence, the filter-based methods can cause an over-smooth problem at the depth discontinuity regions, and also the MRF-based method can occur with error propagation during the optimisation process.

Rendering with explicit geometry technique has been used in sections 4.2 and 4.4 of chapter four of this thesis.

2.3.2 *Rendering without Geometry*

In another IBR method comprising of plenoptic function, light field, mosaic, and concentric mosaic techniques, initial knowledge of the geometry of the scene object is not required. However, a large number of image samples is a significant requirement for good quality rendering [112]. Plenoptic function is the parameterisation which is representative of the intensity of a light ray passing through the camera centre at a 3D spatial location, (x, y, z), in a certain 2D viewing direction, (θ, ϕ), with a certain wavelength, λ, and at a certain time, τ. (Seven-Dimensional (7D)) plenoptic can hardly be realised in reality. It is thought to be an idealised concept. However, it is a platform for the initialising the formulation of other IBR techniques in which no geometry is required [113].

Light field rendering technique is formulated from first principle based the plenoptic function, which summarises all of the radiant energy that can be perceived by an observer at any points in space and time. The (Four-Dimensional (4D)) light field has been established as a promising paradigm to describe the visual appearance of a scene. Compared to a traditional 2D image, it offers information about not only the accumulated intensity at each image point, but also separate intensity values for each ray direction. This means that both the phase and amplitude of the light scattered by scene objects are recorded. Thus, the light field implicitly captures 3D scene geometry and reflectance properties

[20, 114]. Filtering and interpolation are basically the two steps involved in rendering a new image from a set of image samples [115–128]. A wide range of applications have been motivated by the additional information embedded in a light field. Virtual viewpoint creation [60, 129, 130] allows a scene to be rendered with a look different from any of the reference images. As observed in [20], the light field data also allows to add effects like synthetic aperture, i.e., virtual refocusing of the camera, stereoscopic display, and automatic glare reduction as well as object insertion and removal. It needs to be mentioned that data compression is highly required because of the large amount of data.

Lumigraph is a variation of light field rendering in which the knowledge of approximate geometry is key to the achievement of robust and efficient rendering performance [131–136]. Also, concentric mosaic does not require geometric information. However, the camera used in the acquisition of image samples has several centres of projections which trace a circle path [137]. Hence, it is a 3D parameterisation of plenoptic function. It features a considerable reduction in data when compared with light field rendering. A 2D parameterisation of plenoptic function is the mosaic. In the processing leading to the acquisition of multiple image samples of the scene, the camera is constrained to a fixed point and panned. Using feature-based image registration, a motion model can be developed which allows the acquired image samples to be stitched on a cylindrical, spherical, cubic, and polar projection surface [138, 139].

The works presented in Chapter Four under section 4.4 are based on rendering without geometry.

2.3.3 *Rendering with Implicit Geometry*

In literature, a category of IBR exists which can deal with implicit geometry. Implicit geometry involves the establishment of feature correspondence between the participating image samples. The accuracy of this technique depends on application of robust feature detection, extraction, and matching algorithm. As an example, feature correspondence is difficult to establish for a texture-less surface. Popular techniques in this category include view morphing [140–144], transfer method based on trifocal, fundamental matrix or depth geometric constraint [145–150], and view interpolation [151–160]. In view morphing, a new image can be synthesised provided the viewpoint lies along the straight line which joins the centre of projection of the two cameras which provide the reference images. This condition implies that viewpoint translation must be linear and perpendicular to the line-of-sight of the physical cameras. It basically provides the average of two existing parallel views at a viewpoint. The reference images are first rectified if they are not parallel. Transfer method establishes the relationship between coordinates of feature points in multi-views using the concepts of fundamental matrix epipolar or trifocal tensor. These are important tools for understanding the image formation process for several cameras and for designing reconstruction algorithms. In the formulation of view synthesis problem, epipoles play an important role and can be computed from a 3 × 3 fundamental matrix which relate two camera views or a 3 × 3 × 3 trifocal tensor for three views. View interpolation allows arbitrary viewpoint to be created under the condition that the interocular distance of the reference cameras is narrow. This ensures considerable overlap of the reference images in order to minimise occlusion.

2.4 METRIC ASPECT OF IMAGE-BASED RENDERING

This study seeks an understanding of the role played by region match measures in the generation of depth maps which consequently affects the quality of textured depth map with respect to a virtual viewpoint in between two reference images obtained from existing cameras. Virtual camera realisation based on the pixel information contained in left and right image frames is the concern of this section. An important aspect of this implementation is the choice of region match measure type which significantly impacts on the quality of the generated depth map and consequently the rendered image. The region match measures widely used in image processing to solve correspondence problem are Sum of Absolute Differences (SAD), Sum of Square Differences (SSD) and Normalized Cross-Correlation (NCC).

2.4.1 *Introduction*

It is a widely held opinion that only vision provides sufficient information in order for advanced life-forms to infer the correct responses under a variety of circumstances. Perhaps this is well appreciated, perhaps not; there has been a tremendous continuity in the advancement and development of visual media by which man is able to record light and make sense out of it.

3DTVand FTV constitute a convincing evidence of eminent feature which points to the expectation that 3D video will become the choice for visual communication via the internet [2, 11, 25, 161, 162]. Both 3DTV and FTV use a number of time-varying configurations of physical cameras to provide multiple views of the same scene, popularly known as MVV [163].

A consistent position through computer vision research field is that starting with some reference images, new images of a scene can be synthesised as if they were taken from a virtual viewpoint which is different from all the viewpoints of the real images [37, 164]. This is called IBR. It has the significant advantage of reducing the number of physical cameras used in the acquisition of information about the geometry and optical properties of the environment provided the architecture and density of cameras involved have not been compromised.

According to [46, 165, 166], the spectrum of IBR can be classified into three continuum groups namely rendering with explicit geometry, rendering with implicit geometry, and rendering with no geometry.

In the first category the whole 3D structure of the scene is reconstructed. Depth map is used in prominent techniques such as 3D warping as in Figure 2.6, Layered-Depth Images (LDI), and LDI tree. A set of images of a scene and their associated depth maps are used to create a scene model. In surface light field geometry-based IBR representation, images and cyberware scanned range data are the main constituents that are required. When depth is available for every point in an image, the image can be rendered from any nearby point of view by

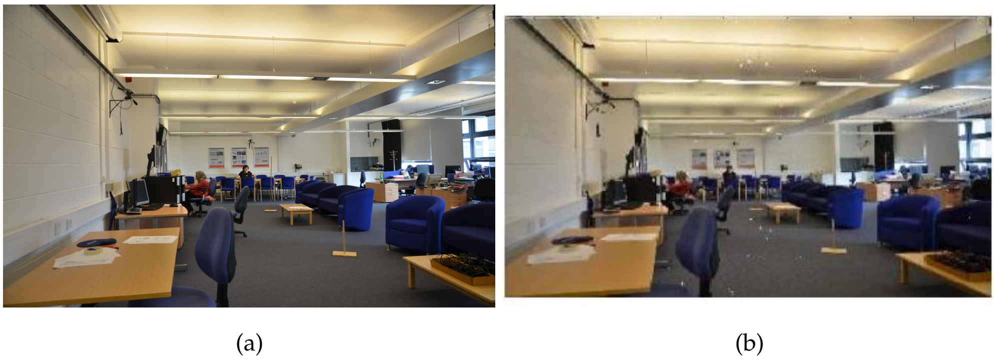

(a) (b)

Figure 2.6: Rendering with explicit geometry. (a) real scene (b) rendered scene using 3D warping.

projecting the pixels of the image to their proper 3D locations and re-projecting them onto a new picture.

References [161, 167] emphasised that the determination of dense stereo matching is limited to estimating an accurate depth map due to the failure of correspondence point matching on the texture less and occluded regions. On the other hand, active depth sensors can only create depths of nearby objects in a lower resolution. However, point correspondence is closely related to feature matching parameters which establish the degree of correlation between two images.

In [168], a constrained energy minimisation problem is formulated. The solution provides image warping functions which are used to render new views. However, it is a generally accepted fact that Depth Image-Based Rendering (DIBR) allows for synthesis of novel views, using the information contained in the depth map, with respect to a virtual view point. It also provides for a reduced bandwidth required for its transmission in 3DTV and multimedia systems applications [161].

One of the objectives of this thesis concerns 3D warping with implicit geometry. We seek to determine the effect of region match metric on feature correspondence between a pair of stereo images on feature matching parameters. This consequently affects the quality of depth map and synthesised images. Providing a three dimensional impression in multimedia applications is a challenging task that requires accurate depth information. Depth map is an image that contains information about the distance, with respect to a view point, to the surface of scene object as a function of image coordinates.

2.4.1.1 *Intensity Image Match Measures*

Finding the pixel coordinates in two different intensity images that correspond to the same point in the world is a data association problem and has been intensively studied in computer vision. Sum of SAD and SSD are two popular computationally inexpensive image region match measures [169]. In (2.2) and (2.1), I_1 and I_2 are compatible image regions with (x, y) and $x, +d_x, y + d_y$ being points in their respective local coordinate spaces. The significant problem with these typical matching cost / aggregation methods is the assumption about the Gaussian noise distribution and, as a consequence, the choice of SAD and SSD is not well justified in many cases of image feature matching since they are susceptible to outliers [170].

$$SAD = \sum_{(i,j)\epsilon U} \left| I_1\left(x+i, y+j\right) - I_2\left(x+d_x+i, y+d_y+j\right) \right|. \tag{2.1}$$

$$SSD = \sum_{(i,j)\epsilon U} \left(I_1\left(x+i, y+j\right) - I_2\left(x+d_x+i, y+d_y+j\right)\right)^2. \tag{2.2}$$

Assuming signal preconditioning, zero mean SAD and SSD, or Covariance Variance (CV), is able to minimise the effect of variation in camera intrinsic parameter and noise perturbation. For applications such as stereo matching or image registration in the log-polar space, CV is a preferable choice. It is invariant to the linear transformation of the two matched image signals.

Sum of Cross Product (SCP) and Normalized Sum of Cross Product (NSCP) constitute another region match measure based on intensity signals of matched blocks of images and scalar product between two vectors. The normalisation condition is not sufficiently satisfied in this case. This is depicted in (2.3) which shows that the mean resultant signal energy in the matched image region equals SSD plus twice SCP. Therefore, it is al-right to use SSD instead of SCP.

$$SSD\left(I_1, I_2\right) = \sum_{(i,j)\epsilon U} \left(I_1^2\left(x+i, y+j\right) + I_2^2\left(x+d_x+i, y+d_y+j\right)\right) \\ - \sum_{(i,j)\epsilon U} SCP\left(I_1, I_2\right). \quad (2.3)$$

There are many recent algorithms aimed at addressing the drawbacks associated with NCC algorithm expressed in (2.4). However, NCC still remains universally recognised [171].

$$\gamma\left(u,v\right) = \frac{\sum_{x,y}\left[f\left(x,y\right) - \bar{f}_{u,v}\left[t\left(x-u,\, y-v\right) - t\right]\right]}{\left\{\sum_{x,y}\left[f\left(x,y\right) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t\left(x-u,\, y-v\right) - t\right]^2\right\}^{0.5}}. \quad (2.4)$$

The work presented in section 4.2 of chapter four is partly based on the discussion in this section.

2.5 GENERATION OF THREE-DIMENSIONAL PANORAMIC IMAGE

It states that three-dimensional content can be generated from two slightly different panoramic views of a scene. The strength of this thought lies in both the intrinsic movement parallax and single effective viewpoint that is associated with the panoramic view of a scene. These factors when effectively put together have the potential to provide pure perspective image. Three-dimensional effect in a single panoramic image is realized by superimposing a pair of panoramic view images. A special fascination of the resultant image is that depending on how the initial panoramic views are generated, the natural scenery and three-dimensional effect in the composed anaglyph panorama differ. A pair of anaglyph glasses would be used by a group of people to view the depth effect in the stereoscopic panorama.

Panoramic mosaic has gained prominence as a 2D form of seven-dimensional plenoptic function widely used for synthetic wide-angle camera. Starting with either an odd or even number of images, the focus is on the generation of panoramic view of a scene with a significant structural difference when compared with the conventional panorama.

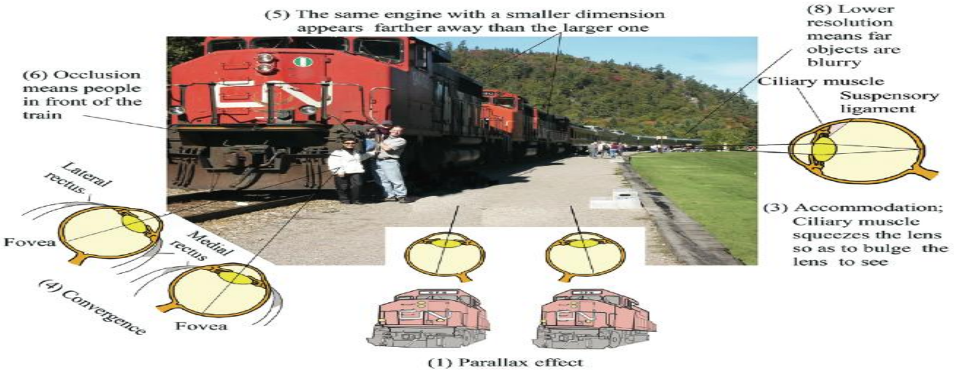

(5) The same engine with a smaller dimension appears farther away than the larger one
(8) Lower resolution means far objects are blurry
(6) Occlusion means people in front of the train
Ciliary muscle
Suspensory ligament
Lateral rectus
Fovea
Medial rectus
(3) Accommodation; Ciliary muscle squeezes the lens so as to bulge the lens to see
(4) Convergence
Fovea
(1) Parallax effect

Figure 2.7: Perspective projection in humans. Taken from [186].

2.5.1 *Introduction*

It is widely understood that a change of viewpoint with respect to an observed object provides either a slightly or considerably different view of the object. This effect which is attributable to the fact that humans have two eyes and see through perspective projection as shown in Figure 2.7 has been extensively studied with regards to HVP and Machine Vision System (MVS). Also, many complex visual tasks, such as reading, detecting camouflaged objects, and eye-hand coordination are also performed more effectively with two eyes than with one, even when the visual display contains no depth [172].

The degree of perceived 3D realism and enhanced Field of View (FOV) are two important factors in vision analysis. In the work of [173], it is observed that retriever of information on the 3D structure and distance of a scene, from a stereo pair of images has become a popular concept in computer vision. A refined analysis has indicated that emerging areas of application in multimedia, with extraordinary standing such as 3DTV and FTV are some of the driving factors for this development [174]. In some relevant applications robustness, accuracy, and real-time capability are of utmost importance as depicted in Figure 2.8.

Multi-view video is one of the enabling technologies which have recently brought 3DTV and FTV to prominence [165, 175]. In spite of the enormous advantages associated with 3DTV and FTV, [176–179] have noted the bandwidth requirement and other issues, which are critical and challenging for transmitting additional data to render the auxiliary view(s).

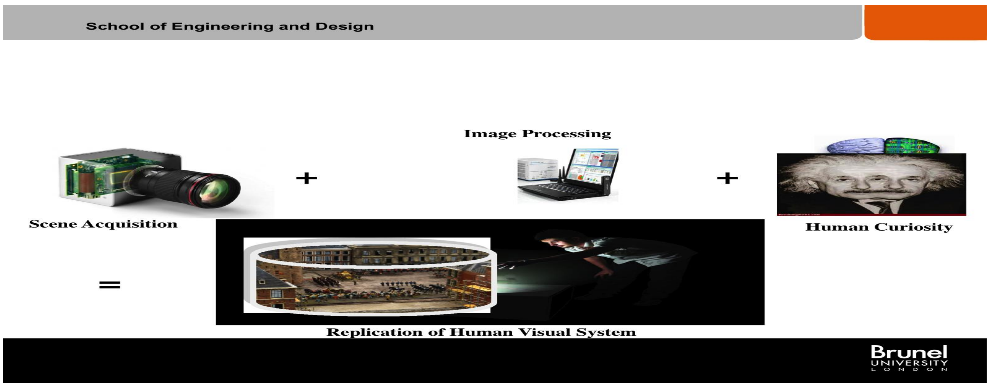


Figure 2.8: Replication of human vision system. Partly taken from [11].

Enhanced FOV is the main motivation factor of [180]. It is emphasized that for any FOV enhancement to be achieved, the entire imaging system must have a single effective viewpoint to enable the generation of pure perspective images from a sensed image. The single viewpoint constraint is satisfied by incorporating reflecting mirrors into the conventional imaging system.

The generation of 3D content from a stereo pair of panoramic images of a scene is possible. In the view of [181], the following advantages cannot be divorced from stereoscopic view. Depth perception relative to the display surface; spatial localization, allowing concentration on different depth planes; perception of structure in visually complex scenes; improved perception of surface curvature; improved motion judgement; improved perception of surface material type. These benefits give stereoscopic displays improved representation capabilities that allow the user a better understanding or appreciation of the visual information presented.

Panoramic images have been widely investigated in the work of [182–184]. It is also a variant of image-based rendering that allows 3D scenes and objects to be visualized in a realistic way without full 3D model reconstruction. The concept of panoramic image stitching stems from the fundamental deficit in the narrow field of view FOV of most compact cameras as depicted in Figure 2.9.

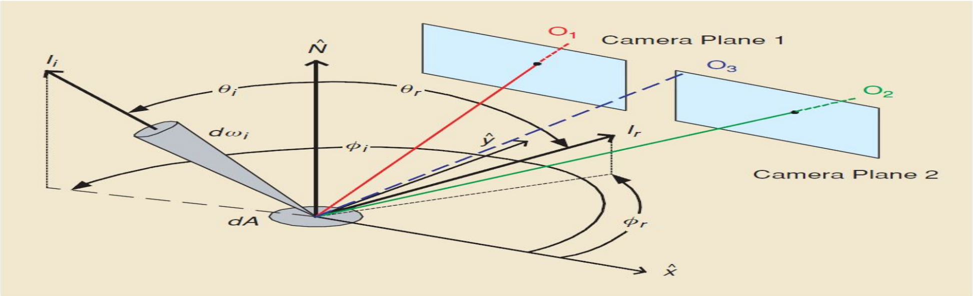


Figure 2.9: A compact camera image formation process. Taken from [165].

2.5.2 *Three-Dimensional Visual Content*

2.5.2.1 *Binocular Vision and Stereoscopy*

Binocular vision involves the use of two eyes or optical devices for the acquisition of both the optical and geometric properties of a scene. It is thought to provide for increased field of view, binocular summation which is the enhancement in the detection of faint objects, and the use of stereoscopic distance or disparity to perceive a scene in 3D and the distance of an object [185]. The amazing effect of significant proportion is the composition of a single image using the single individual image of each eye. This is generally referred to as binocular fusion. The superposition of a pair of images to create depth illusion is known as anaglyph.

In [186], it is believed that parallax, movement parallax, accommodation, convergence, remembered geometry of an object and linear perspective, occlusion, shading, and resolution constitute both physiological and psychological factors, which determine the level of 3D effect we observe as humans. However, parallax and convergence are the most needed factors for anyone to perceive 3D effect. With accommodation, neurophysiological process varies the radius of curvature of the eye lens to focus the image on the retina. However, with convergence, the continuous movement of the eye ball causes certain angle which decreases with distance to be subtended between the visual axis and optical axis of each eye. This is perhaps linked to the availability of neural algorithm which plays a prominent role in the binarization and manipulation of information the eyes receive.

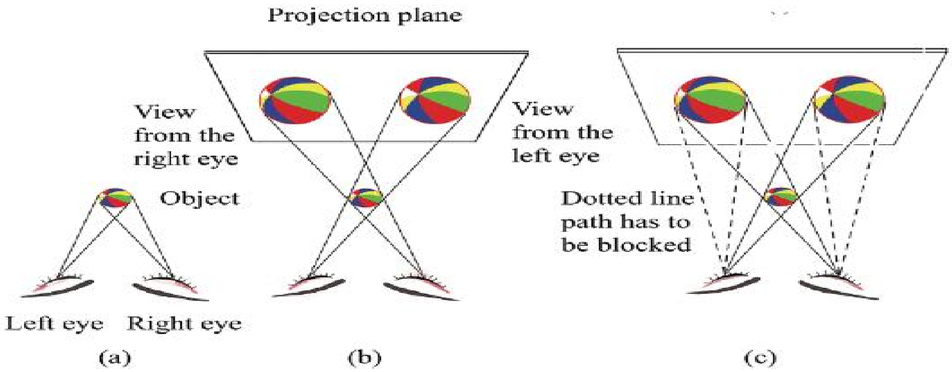


Figure 2.10: Parallax effect. (a) Object. (b) Projected views of the object. (c) Transposed images. Taken from [186].

2.5.2.2 *Anaglyph and Synthetic 3D Effect*

At man-made level, the singleness of vision created by neural algorithm in humans is reversed. There are several stereoscopic display methods that can be used to generate 3D effect. These include lenticular sheet, integral photography, horse blinder barrier, parallax barrier, varifocal mirror, volumetric methods, head mounted display, time sharing method, anaglyph, Brewster's stereoscope, Wheatstone's stereoscope, and 3D movies [186]. From either the projection or interception type of display, one of the two slightly different images of the same object captured with two similar cameras separated by a certain stereoscopic distance is presented to each eye alternately through a filter glass. This concept is demonstrated in Figure 2.10.

Whatever display type is used, comfortable view in terms of reduced eye strain or absence of double images from excessive perceived depth is highly required [187, 188]. In [181], it is stated that the mentioned requirement is a function of stereoscopic camera parameters. It is further mentioned that a stereoscopic camera system with parallel axes should be used to avoid the vertical image disparity generated by systems that verge the camera axes. This is because for a parallel camera system, points at infinity have zero disparity and are perceived by the viewer in the plane of the target display. To ensure that corresponding points in the left and right images, at other distances from the viewer, are perceived in the screen plane, the images must be adjusted during or after capture. All these explain the difficulty in producing comfortable images which are often only produced after repeated trial and error. Some common challenges are are highlighted in Figure 2.11, [42].

2.5.2.3 *Anaglyph Mathematics*

In **??**, two similar cameras with a focal length of f having a stereoscopic distance of b between them are used to acquire a world point (X, Y, Z). The relationship between the world point and the respective corresponding points (x_R, y_R) in the right image and (x_L, y_L) in the left is expressed as

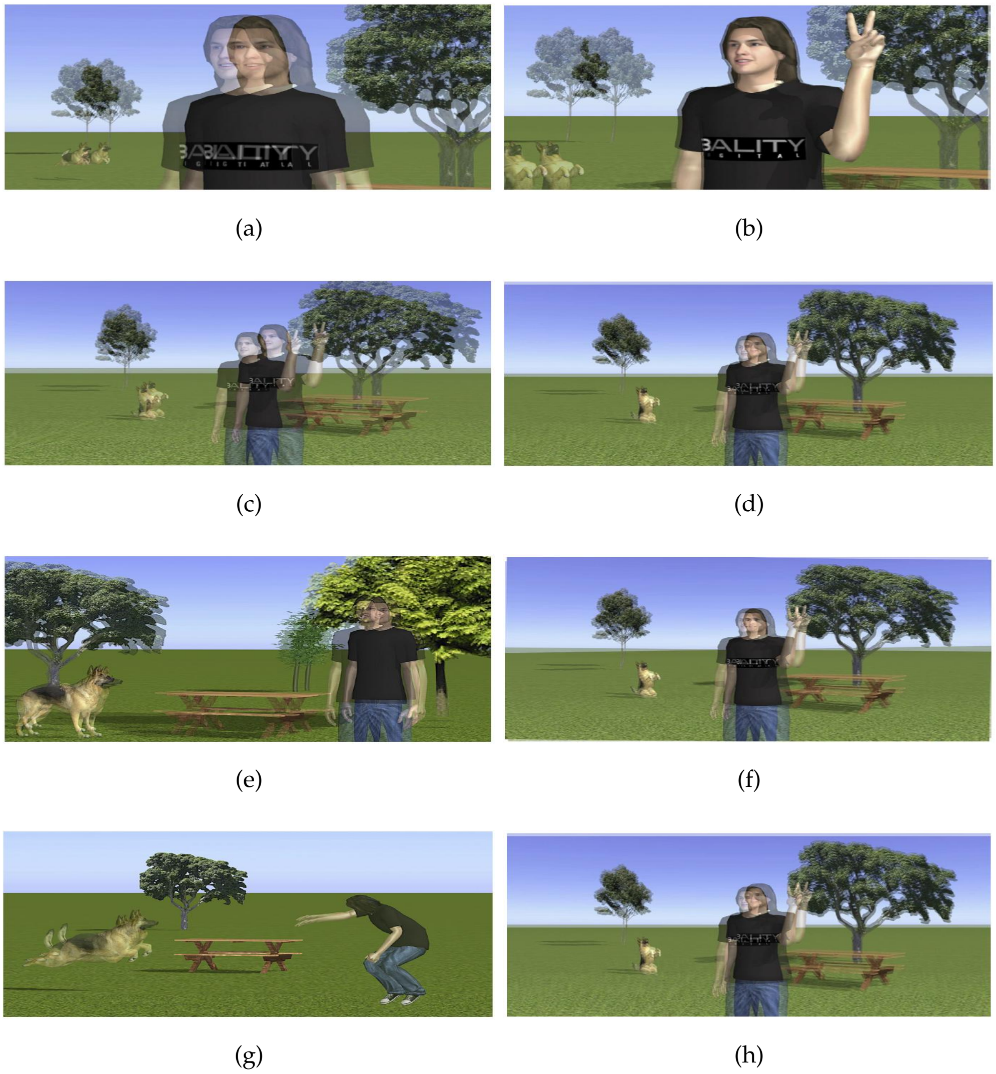
$$\frac{x_L}{f} = \frac{x + \frac{b}{2}}{z}, \frac{x_R}{f} = \frac{x - \frac{b}{2}}{z}, \frac{y_L}{f} = \frac{y_R}{f} = \frac{Y}{Z}. \tag{2.5}$$

The disparity between corresponding right and left image points is expressed in (2.6). The reciprocal of disparity gives the depth of the world point with respect to the vertical plane containing the cameras and it decreases with stereoscopic distance. It is also important to note from (2.5) that disparity D, is directly proportional to the product of camera focal length and stereoscopic distance, and inversely proportional to the depth.

$$D = x_L - x_R. \tag{2.6}$$

It has now been proven that a convincing and comfortable viewing experience can be realized not by maintaining a certain angular disparity as earlier suggested by human factor studies [189] but by compression of scene depth. In [181], this idea is depicted as shown in Figure 2.13.

In the simplified case of a static viewer analysed in [181], camera separation b can be computed using the relation (2.7). Where $Z^{'}$ is the distance of the cameras from the 'virtual' display (zero-Disparity-Plane) in the scene, $N^{'}$ is the distance from the cameras to the closest visible points in the scene, d_N is the disparity, on the display, of objects appearing at the limit N.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 2.11: Causes of 3D discomfort: (a) the distance between the cameras is not adequate, (b) cameras were converged on the wrong point, or one eye was excessively horizontally shifted, (c) lens mismatch, (d) poor manual control on focus distance, and autofocus modes may disagree on the subject distance, (e) keystone appears when the optical axes are not parallel, due to convergence or, less often, strong vertical misalignment, (f) image rotation appears when the camera's optical axis is rotated along the Z axis, (g) both left and right images are shot without paying great attention to time synchronisation, (h) one camera is most likely pointing up or down, or a zoom is off-axis. Taken from [42].

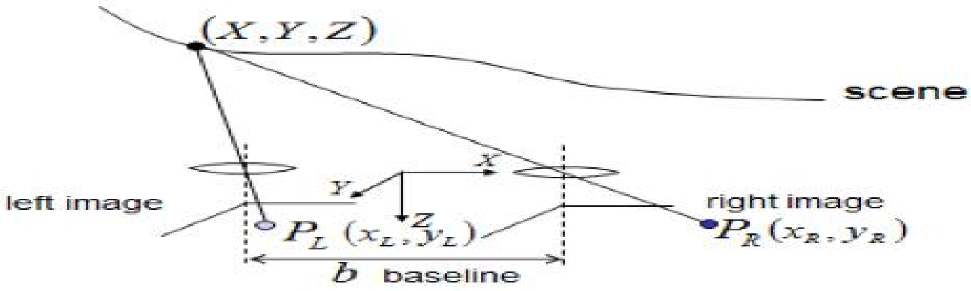


Figure 2.12: A compact camera image formation process.

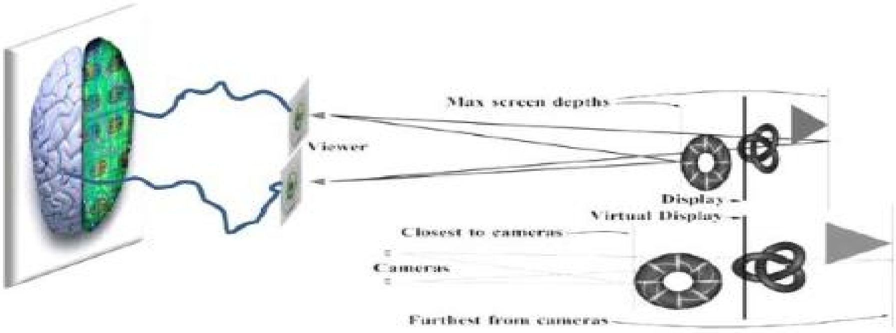
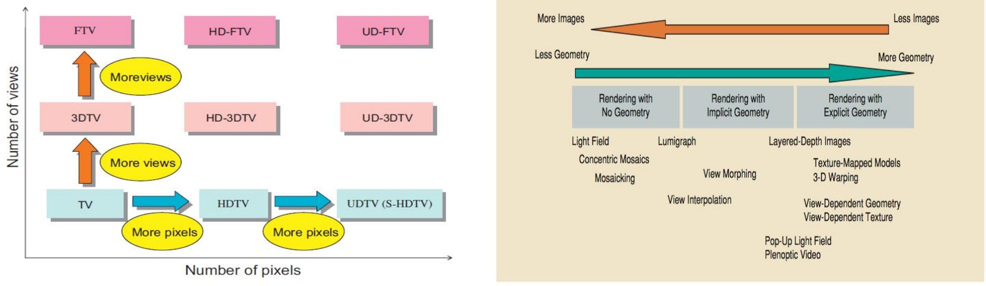


Figure 2.13: The scene depth (bottom) is compressed (top). Taken from [181].

$$b = \frac{2Z' \tan\left(\frac{\theta}{2}\right) d_N N'}{W\left(Z' - N'\right) + d_N N'}. \tag{2.7}$$

The following five types of anaglyph are well known in computer vision. True anaglyphs, colour anaglyphs, grey anaglyphs, half colour anaglyphs, and optimised anaglyphs. According to [190], colour is the general name for all sensations arising from the activity of the retina of the eye and its attached nervous mechanisms, this activity being, in nearly every case in the normal individual, a specific response to radiant energy of certain wavelengths and intensities. This understanding can be explored to seek a mathematical representation of anaglyph. In terms of implementation, colour and grey anaglyphs are usually composed based on the mathematics expressed in (2.8) and (2.9) respectively. A_r, A_g, A_b are the colour components of the anaglyph generated from panoramic views 1 and 2 with r, g, b colour components.



(a) (b)

Figure 2.14: (a) categorisation of television, (b) categorisation of image-based rendering. Taken from [2] and [165] respectively.

$$\begin{bmatrix} A_r \\ A_g \\ A_b \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1_r \\ 1_g \\ 1_b \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2_r \\ 2_g \\ 2_b \end{bmatrix}, \tag{2.8}$$

$$\begin{bmatrix} A_r \\ A_g \\ A_b \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1_r \\ 1_g \\ 1_b \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0.299 & 0.587 & 0.114 \\ 0.299 & 0.587 & 0.114 \end{bmatrix} \begin{bmatrix} 2_r \\ 2_g \\ 2_b \end{bmatrix}. \tag{2.9}$$

2.5.3 *Panoramic Image*

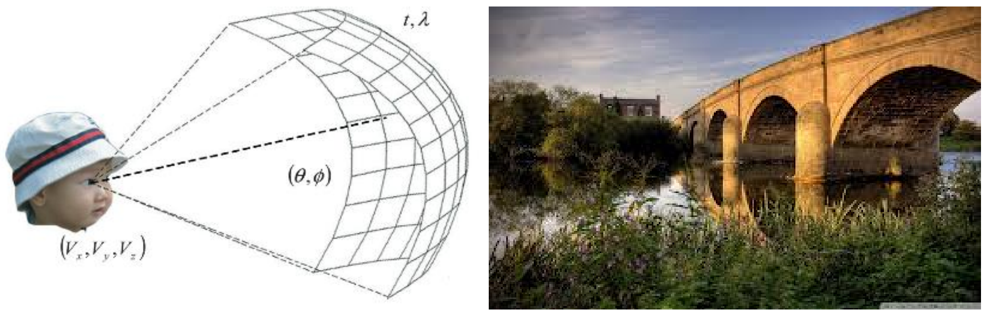
Multi-view video [2] is one of the enabling technologies which have recently brought 3DTV and FTV to prominence as is shown in Figure 2.14 (a). The subject of intensive research has been to optimise the architecture while minimizing the number of physical cameras used in the acquisition of both the physical and geometric properties of a scene [165, 175]. Virtual cameras can provide an excellent minimization of physical cameras through the concept of IBR. In IBR, a completely distinct description of the world recorded in the reference images is explored. It is popularly categorized as shown in Figure 2.14 (b).

According to [53, 60, 113, 191–193], plenoptic function is sufficiently representative of the pattern of dense array of light rays filling the environment, Figure 2.15 (a). Therefore, any image acquired using an optical system can be considered as a sample of the plenoptic function which is continuous in nature. Concisely explained, plenoptic function expresses the intensity i, of light rays passing through the camera centre at every location $\left(V_x, V_y, V_z\right)$ at every possible angle (θ, ϕ), for every wavelength λ, at every time t. This is expressed in (2.10) and depicted in Figure 2.15 (a). As an example, a sample of plenoptic function which represents a perspective projection of a scene is shown in Figure 2.15 (b).

It is stated in [59, 194] that IBR techniques and algorithms can be regarded as the result of plenoptic function dimension reduction. For example, a panoramic view on the fixed viewpoint is a two dimensional reduction of plenoptic function. Concentric mosaic is a three dimensional reduction. Light Fields and Lumigraph reduce the plenoptic function to four dimensions. These techniques can create novel views of scenes or objects by resampling a database of images and representing a discrete sample of the plenoptic function. A concise problem statement for IBR paradigms, such as morphing and view interpolation is provided in [60]. A reduction to a five-dimensional function varying with the viewpoint and the viewing direction is the main focus. However, surface plenoptic function is a variation of six dimensions.

IBR techniques at one end of the spectrum of Figure 2.7 (b) such as lightfield, lumigraph, concentric mosaics, and mosaicking rely on dense image sampling of the scene and remain attractive in the conversing computer vision industry. The attractiveness as noted in [165] is on the merit of no or very little geometry information for rendering without recovering the exact 3D models, superior image quality, and a considerably less computational resources for rendering regardless of the scene complexity, because most of the quantities involved are precomputed or recorded. Furthermore, the implementation of these techniques is enhanced through change of viewpoints and sometimes limited amount of relighting [195].

Panoramic photography is the most thoroughly discussed IBR technique which requires no 3D model of the scene [196]. In the view of [197], the strength



(a)

(b)

Figure 2.15: (a) 7D plenoptic function, taken from http://www.google.co.uk/search?q=plenoptic+function (b) perspective projection.

of this IBR method stems from the existence of a simple invertible transformation between images gathered with a camera rotating around its centre of projection.

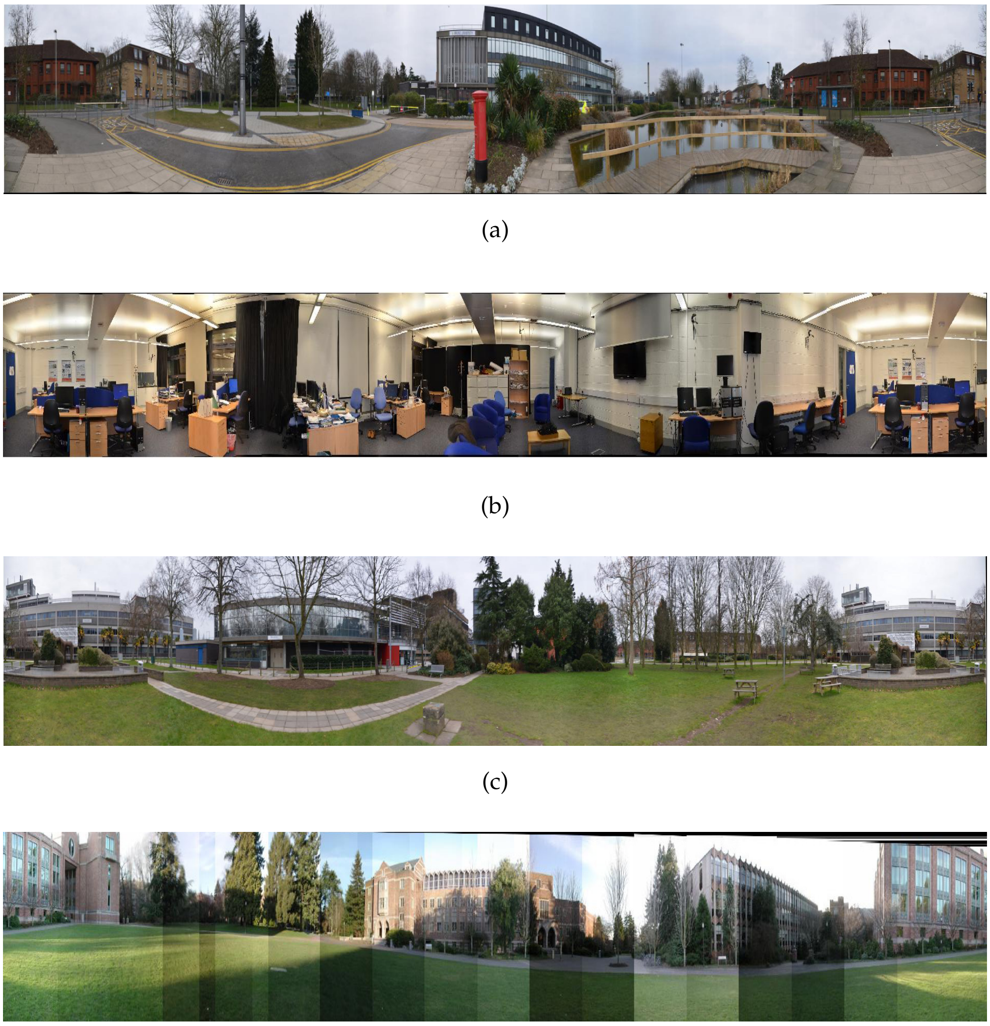
The usual choice for compositing larger panoramas is to use a cylindrical or spherical projection in order to avoid excessive stretching of pixels near the border of the image. However, the potential presence of large amounts of independent motion, camera zoom, synthesis of virtual environments, and the desire to visualize dynamic events impose additional challenges.

A unique challenge which has been identified by [184] to play an important role in the construction of panoramic image is the radius of the projection surface. Distortion minimization has been the sole argument in favour of the equality of camera focal length and radius of projection surface also known as scale factor. This is a fundamental assumption which is deeply entrenched in most of the previous works in which cylindrical warping has been used as part of their implementation strategy.

Inspired by the complexity of seven-dimensional 7D function in (2.10), a classification can be considered to obtain its 3D version called concentric mosaic.

$$i = I\left(V_x, V_y, V_z, \theta, \phi, \lambda, t\right) \tag{2.10}$$

This will be made possible if there is a simplification of light ray wavelength to red, green, and blue channels, constancy of radiance along a light ray through empty and transparent space, absence of time dimension, occurrence of 2D mo-



(a)

(b)

(c)

(d)

Figure 2.16: A set of conventional panoramic images. (a) Brunel Pond, (b) Cmcr, (c) Hsbc, (d) obtained from www.cs.washington.edu.

tion which restricts both the camera and the scene to a plane [182, 198]. Interestingly, by imposing the most restrictive condition of fixed viewpoint, that is, no motion of viewpoint, a specialised concentric mosaic and a 2D plenoptic function called image mosaic is realized. Image mosaicking actually involves the composition of one single mosaic with multiple input images representative of a static scene and acquired by a panning camera. A few examples of 2D plenoptic function generated during the course of this research are shown in Figure 2.16.

Both Image mosaicking with SCOP also known as panoramic image and multiple centre-of projection type are characterized by increased field of view angle and light rays are indexed by their directions (θ, ϕ). This provides a unique formulation and understanding which allows for the construction of a feature-based technique for aligning frames previously warped into a spherical or cylindrical surface. A special fascination of a panoramic mosaic which has been emphasized by expert studies is its application for increasing the resolution of images termed as super-resolution, object insertion, and texture synthesis [183, 184].

The calibration and understanding of the enormous challenges which are usually encountered during the implementation of panoramic mosaic demand to be mentioned. The use of multiple images in a panoramic mosaicking demands that certain control techniques be adopted to compute a globally consistent set of alignments and to efficiently discover which images overlap one another. For 360 degree field of view, an understanding of the final compositing surface namely: a cylinder or sphere, onto which to warp and place all of the aligned images is absolutely essential. We also need to develop algorithms to seamlessly blend overlapping images, even in the presence of parallax, lens distortion, scene motion, and exposure differences [184].

A problem of significant dimension in the domain of panoramic mosaicking which is linked to feature point matching has been noted by [199]. It is observed that perfect scale invariance cannot be achieved in practice because of sampling artefacts, noise in the image data, and the fact that the computational effort limits the number of analysed scale space images. For any object in an image, interesting points on the object can be extracted to provide a feature description of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

In the work of [200], the emphasis is on the construction and evaluation of local descriptor representations. The line of reasoning points to the fact that although localization and description aspects of interest point algorithms are often designed together, the solutions to these two problems are independent. Another important characteristic of these features is that the relative positions between them in the original scene should not change from one image to another. Feature positions should not depend on the orientation of the object. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed.

As part of the contribution of this research, the potential of panoramic view which is brought into lime light is the concept of perspective projection. When the human eye looks at a scene, objects in the distance appear smaller than objects close by. Perspective definition shows distant objects as smaller to provide additional realism.

2.5.4 *Perspective Projection*

2.5.4.1 *Camera Perspective*

As opposed to orthogonal projection, perspective projection allows objects in the distance to appear smaller than objects close by to provide additional realism. This is the way we see as humans. Reference [201] confirms that real cameras exist with more complex full parameterisations aimed at implementation of the effects of perspective projection. Perspective projection performs a mapping from 3D space to the two-dimensional image plane during which straight lines in the world are projected to straight lines on the image plane. Parallel lines in the world are projected to lines that intersect at a vanishing point. In drawing, this effect is known as foreshortening. The exception is lines in the plane parallel to the image plane which does not converge.

Furthermore, conics in the world are projected to conics on the image plane. For example, a circle is projected as a circle or an ellipse. A characteristic which is noteworthy is that mapping is not one-to-one and a unique inverse does not exist. That is, given (x, y) in image plane, we cannot uniquely determine (X, Y, Z)

in the world. All that can be said is that the world point lies somewhere along the projecting ray. [190], observes that the transformation is not conformal – it does not preserve shape since internal angles are not preserved. Translation, rotation and scaling are examples of conformal transformations. A general affine transformation comprises translation, rotation and different scaling for each axis and is not conformal.

The extrinsic and intrinsic parameters of a camera allow for transformation from world frame to camera frame and from image frame to sensor frame respectively. The combination of these parameters constitutes the model of a camera as shown in Figure 2.17. In the presence of barrel distortion, the accuracy and robustness of images to be used in panoramic mosaicking are dependent on this well-understood camera model. Figure 2.17 also shows the transformations involved in the camera model shown in (2.11).

$$\tilde{p} = \begin{bmatrix} \frac{f}{\rho_w} & 0 & u_o \\ 0 & \frac{f}{\rho_h} & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} (T_C)^{-1}\, \tilde{P} = KP_oT_C^{-1}\tilde{P} = C\tilde{P}. \qquad (2.11)$$

Perspective images have a particular relevance of making subsequent processing easier using some of the well-established computer vision techniques which assume perspective projection. Perspective projection exhibit less distortion compared to other forms of projection. Of course, this is good for the human eye.

2.5.4.2 *Image Mosaic*

The limited field of view of conventional imaging devices such as pinhole camera is a problem which is familiar to computer vision researchers and diagnosed by [202, 203]. It is pointed out that while surveillance, teleconferencing, and model acquisition for virtual reality constitute a driving force for an increased field of view, there are several other application areas which are strategically positioned to take advantage of field of view enhancement.

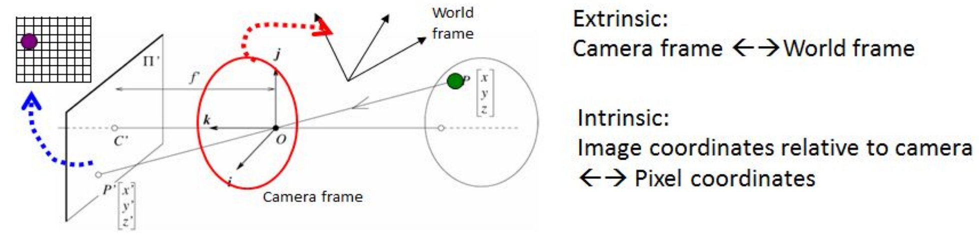


Figure 2.17: Perspective camera: Camera model and Transformation between world frame and sensor frame.

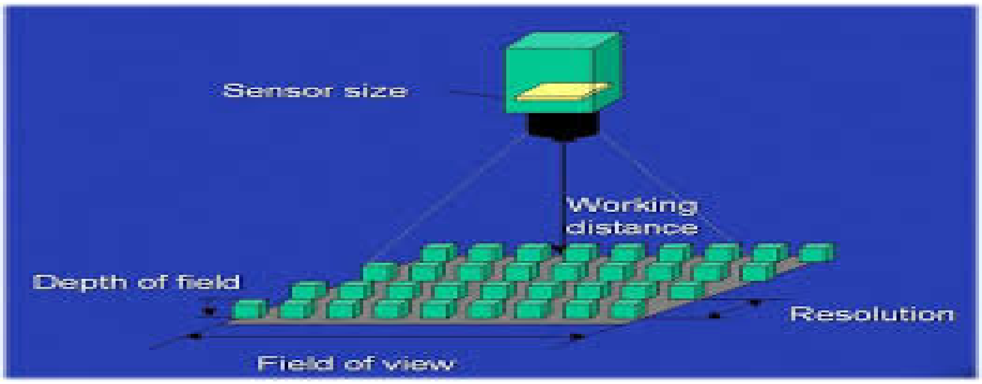


Figure 2.18: Field of view and resolution of a camera. Taken from http://www.google.co.uk/search?q=field+of+view.

Catadioptric image formation process is widely used for enhancing the field of view of imaging devices. However, image mosaic is favoured since catadioptric image formation has associated problems of sensor resolution and focusing as depicted in Figure 2.18.

Furthermore, in the thoroughly thought-through gradient domain approach, [204], of image stitching, the similarity of the sample images and visibility of the seam constitute the cost functions to be optimised. This eventually suppresses both photometric inconsistencies and geometric misalignments between the stitched images. In image mosaicking, images are first projected onto a curved surface using cylindrical, spherical, cubic or polar projection.

The method for the generation of panorama falls into two categories namely direct and feature-based methods. It is clear from [205] that accuracy of image registration and closed initialization are the main differences between the two.

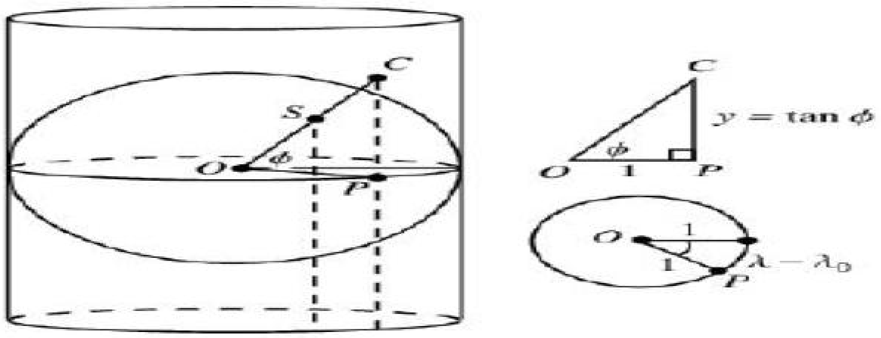


Figure 2.19: Cylindrical Projection.Taken from [206].

Feature-based method is considered in this work, since panoramic view generation is one of the image-based rendering methods and features can only be obtained from reference images.

Of course, the image formation model based on Snell's law developed by [203] is known and well appreciated. In this model, one object point P_0 is traceable to obtain two image coordinates $[u,v]^T$ and $\left[u^{'},v^{'}\right]^T$ on a Charge Coupled Device (CCD) camera by use of skew ray tracing and taking a single-camera two panoramic views. This camera type is yet to be readily available in the market. In this regard, the use of two separate cameras on a single tripod separated by certain stereoscopic distance is inevitable.

In image mosaicking, the image is first mapped onto the surface of a cylinder, sphere, cube and then the curved surface is unrolled. A method to estimate surface projection is well documented in the work of [184]. Cylindrical warping, [206], can be obtained using either forward or inverse warping as depicted in Figure 2.19 and Figure 2.20. In forward warping: From image coordinate (x,y) the projected coordinates on the cylinder $(x^{'},y^{'})$ are given in (2.12) and (2.13) where S is a scaling factor and f equals lens focal length in pixels.

$$x^{'} = S\theta = S\tan^{-1}\left(\frac{x}{f}\right) \tag{2.12}$$

$$y^{'} = Sh = S\left(\frac{y}{\sqrt{(x^2+f^2)}}\right). \tag{2.13}$$

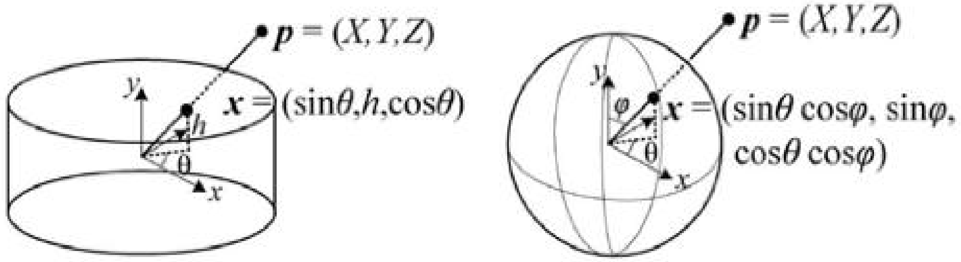


Figure 2.20: Cylindrical and spherical projection.

For inverse warping: inverse mapping from cylindrical coordinates to image (x, y) is expressed in (2.14) and (2.15).

$$x = f\tan\theta = f\tan\left(\frac{x^{'}}{S}\right) \tag{2.14}$$

$$y = h\sqrt{(x^2 + f^2)} = f\left(\frac{y^{'}}{S}\right)\sec\left(\frac{x^{'}}{S}\right). \tag{2.15}$$

For a spherical projection surface, the coordinates have the following relationship for forward warping

$$x^{'} = S\theta = Stan^{-1}\left(\frac{x}{f}\right) \tag{2.16}$$

$$y^{'} = S\varphi = S\left(\frac{y}{\sqrt{(x^2 + f^2)}}\right). \tag{2.17}$$

where (θ, φ) is the parameterised coordinate of a sphere

$$x = f\tan\theta = f\tan\left(\frac{x^{'}}{S}\right) \tag{2.18}$$

$$y = (\tan\theta)\sqrt{(x^2 + f^2)} = f\left(\frac{y^{'}}{S}\right)\sec\left(\frac{x^{'}}{S}\right). \tag{2.19}$$

In conclusion, a discussion on the image capturing techniques, processing, and manipulation of multiple images which lead to the realisation of 3D panoramic image has been presented. In this work, a cylindrical surface whose radius is equal to the focal length value of the acquisition camera will be favoured in the stitching of multiple image samples. Depth perception in the resultant panoramic image is a possibility through anaglyph. Anaglyph can be implemented by superposition of the two previously generated panoramas through stitching process. The ideas presented in this section will take a centre stage in chapter four, section 4.3.

2.6 COMPUTATION OF VIRTUAL ENVIRONMENT FROM STEREO-PANORAMIC IMAGE

Depth map texturing is one of the prominent image-based rendering methods. It requires the condition of epipolar constraint to be satisfied. However, the cylindrical warping involved in the construction of panoramic view makes it highly challenging for this condition to be strictly met. This section reviews the literature related to depth image-based synthesis of virtual environment starting with stereo-panoramic views of a scene.

2.6.1 *Introduction*

The understanding of the phenomena of light reflection and refraction, now fundamental concepts in modern geometric optics have significantly contributed to the discovery of the first camera. Ever since then the computation of both optical and geometric properties associated with 3D structure of scenes and objects within the environment has become the area of influence of cameras based on the pinhole model. Camera has revolutionized the world on the merit of its ability to record light for future use.

Humans are very much familiar with monocular visual cue such as texture gradient, occlusion, and shading through which a meaningful understanding and interpretation of the scene is composed. However, it is the binocular visual cue that has provoked serious intellectual thought. At the inception and conception of the study of optics, it is believed that humans and most predator animals are able to maintain the top position of their different food chains because they possess two eyes (binocular vision) and as a result of which perception of depth is realized through retinal disparity, eye convergence, and parallax. Sir Charles

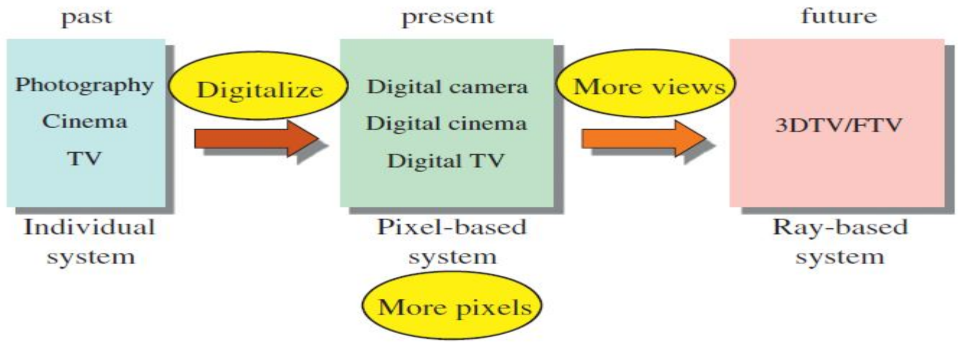


Figure 2.21: Visual media and its development. Taken from [2].

Wheatstone, in 1838, refined this curiosity further. It is stated that the mind perceives an object of three dimension by means of the two dissimilar picture projected by it on the two retinae.

Almost always, light is documented and reconstructed for the purpose of achieving a satisfactory viewing experience. In [2], it is stated that visual media have been through three stages of development namely: Individual systems, pixel-based system, and ray-based system as can be seen in Figure 2.21. For this reason, most cameras have to mimic the computational investigation of the human eyes which are controlled by a wealth of neural algorithms. Robot navigation and object recognition are just a few stocks of areas of application which, partly own their success to camera performance and sophistication. This has been noted to be a significant driving factor responsible for the continuous and active development of camera technology.

The increasing complexity and sophistication embedded in certain emerging areas of application [207–209], which are expected to find a predictably receptive acceptance in the consumer market, have been predicted. The realization of such areas of application will be hugely challenged by a widely held opinion with regards to the narrow field of view FOV of most compact cameras, camera calibration and synchronization, and camera architecture. Having these challenges in sight demands that a procedural and accurate view synthesis be pursued.

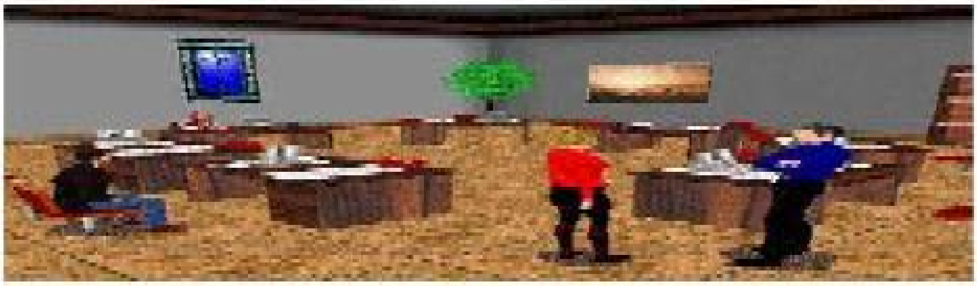


Figure 2.22: Representation of an office scene for collaborative virtual environment (CVE) applications (from German KICK project. Taken from [212].

The 3D driven future development of visual media characterised with photo-realistic visualization potential has remained attractive over the years [165]. The creation of synthetic 3D content via image processing has tremendous appeal since it is possible to use images of real world scenes which usually have more rich details than those generated by using computer graphics techniques [210, 211]. Model-based methods for content generation are less favoured in certain critical application circumstances.

It is known that virtual reality depicted in Figure 2.22 can provide an enhanced distributed collaborative work with group awareness and spontaneous communication capabilities very similar to face-to-face working condition [212]. Virtual reality composed based on panoramic view has the special fascination of realizing the human dream creating a virtual world. Also, consumer on-demand content is fast becoming popular. Panoramic view synthesis is one of the IBR techniques. IBR is composed based on the theoretical concept of plenoptic function [60]. It is a parameterised function for describing everything that is visible from a given point in space. According to this function, (2.10), lighting information can be acquired by many cameras placed in different positions. A panoramic image is a 2D form of plenoptic function.

Therefore, the discussion in this section will be put to use in the synthesis of virtual environment based on depth image obtained from stereo-panoramic views of a scene. The distortion associated with cylindrical warping is a problem. The reason is that the rank-2 constraint is difficult to satisfy. Epipolar constraint is significantly exercised by normalizing the matching points, optimising the intrinsic parameters of the camera and finally, applying singular value decomposition algorithm to the fundamental matrix. This process allows the epipole of one camera to be located in the other and hence making the rectification and other pipeline processes realisable.

2.6.2 *Survey of Imaging Systems*

A large number of compact cameras are based on perspective projection model with a single centre of projection. They are generally characterized by a typically narrow FOV of $50^o \times 35^o$ and hence sometimes called directional cameras. FOV makes compact cameras less popular in certain vision applications. A compact camera is widely used in the acquisition of both static and dynamic scenes. Traditional cameras capture light onto photographic film. Digital cameras use electronics, usually a CCD to store digital images in computer memory inside the camera.

Enhancement of FOV is the main reason for the invention of catadioptric imaging systems. A merit of single view point [180] is significant in the face of the complexity arising from the shape, position, and orientation of the reflecting surfaces employed.

Large FOV above 180^o is the main characteristic of omnidirectional cameras. Furthermore, as pointed out in [213], images of all scene points cannot be represented by intersections of camera rays with a single image plane. For that reason, rays of the image are represented as a set of unit vectors in 3D such that one vector corresponds just to one image of a scene point. Omnidirectional cameras usually find application in areas such as structure from motion, and surveillance where considerably stable ergo-motion computation is expected.

It is stated in [203] that single-camera panoramic stereo imaging systems using an unusual arrangement of curved mirrors, Figure 2.23, is becoming attractive. Such systems solve the problem of capturing a panoramic image with a single motionless camera, and are also capable of positioning and measuring an object with a single imaging process.

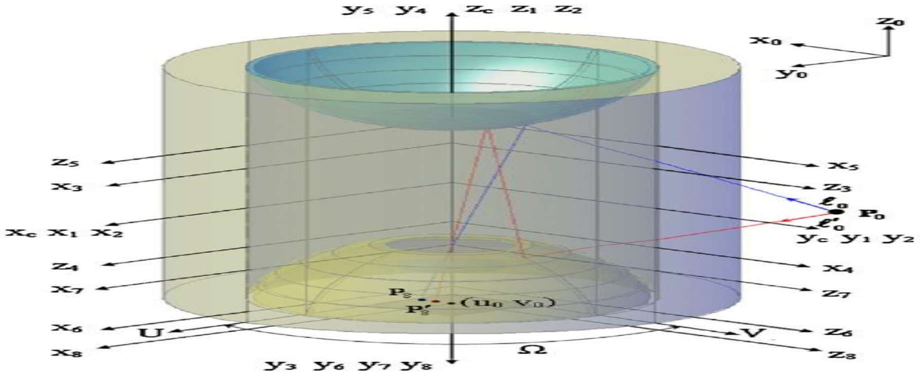


Figure 2.23: Skew ray tracing through a single-camera panoramic stereo system with 8 boundary surfaces. Taken from [203].

2.6.3 *Construction of Panoramic Image*

At implementation level, a panoramic view is the product of computational photography realized through software. It has the established reputation of providing a $360^o \times 180^o$ view of a scene by stitching a set of several images obtained through camera panning through an angle of 360^o. Each of the multiple images contains both the optical and geometric properties of different parts of the scene.

One of the well-known implementation stages of a panoramic view is the warping or projection of each image sample in a set onto a cylindrical surface. A cylindrical radius with a magnitude size which is equal to the camera focal length provides less pixel distortion.

Large number of image samples usually with considerable overlap are stitched together using image mosaicking. Visible seams due to changes of scene illumination and camera responses, or spatial alignment errors are the challenges that require critical attention. Alpha and gradient domain image blending are two transition smoothing approaches to reduce colour differences between source images to make seams invisible and remove stitching artifacts [214].

While alpha blending cannot avoid ghosting problems caused by object motion and small spatial alignment errors, algorithms which implement gradient domain approach [204] can reduce colour differences and smooth colour transitions using gradient domain operations, producing high-quality composite images.

Optimal seam finding approaches search for seams in over-lapping areas along paths where differences between source images are minimal. The seams can be used to label each output image pixel with the input image that should contribute to it.

An interesting quality performance has been reported in [215] where a combination of transition smoothing and optimal seam finding approaches have been used in panoramic view synthesis.

2.6.4 *Generation of Depth Map*

Depth map refers to an image in which is embedded subtle information relating to the distance of surfaces of scenes objects from a viewpoint. It has received enormous investigative attention. Its significant application areas include subsurface scattering, simulating shallow depth of field, and shadow mapping. However, in computer vision, it has enjoyed a more focused adaptation in order to provide the distance information needed to create and generate auto stereograms and in other related applications intended to create the illusion of 3D viewing through stereoscopy [216].

Going by the refined reasoning of [161], a depth map is essentially a 2D function that gives the depth, with respect to the camera position, of a point in the visual scene as a function of the image coordinates. Its realisation is dependent on an important step referred to as image rectification. Image rectification simplifies the search for correspondence points from 2D to 1D problem. An example of rectified image pair is shown in Figure 2.24. Since the depth of every point in an original image is known, a virtual image of any nearby viewpoint can be rendered by projecting the pixels of the original image to their proper 3D locations and re-projecting them onto the virtual image plane. Thus, DIBR permits the creation of novel images, using information from the depth maps, as if they were captured with a camera from different viewpoints.

High quality depth map generation has been thoroughly studied by [161]. An important aspect of this proposition is that a depth map is pre-processed using an asymmetric filter to smoothen the sharp changes in depth at object

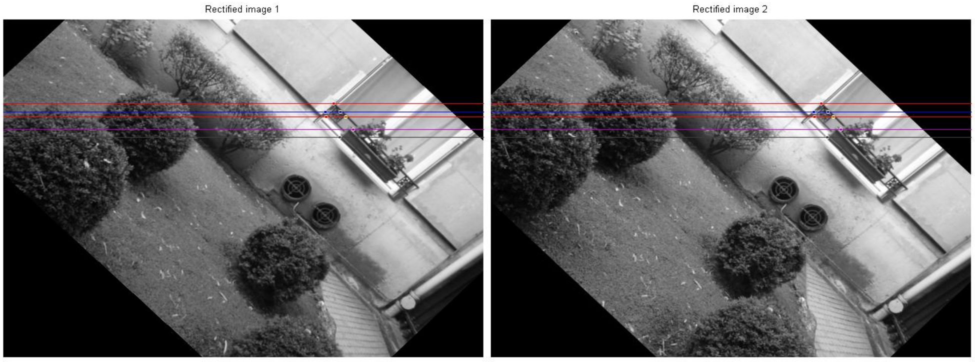


Figure 2.24: Rectified image pair.

boundaries. In addition to ameliorating the effects of blocky artifacts and other distortions contained in the depth maps, the smoothing reduces or completely removes newly exposed (dis-occlusion) areas where potential artifacts can arise from image warping which is needed to generate images from new viewpoints.

In [217], hybrid camera system composed of one Standard-Definition (SD) depth camera and five HD video. The initial depth map is refined using segment-based stereo matching. To reduce mismatched depth values along object boundaries, moving objects are detected using colour difference between frames and extract occlusion and dis-occlusion areas with the initial depth information. Considerations in [218] particularly focus on aspects of inter-operability and multi-view adaptation for the case that different multi-baseline geometries are used for multi-view capturing and 3D display. Furthermore, it presents algorithmic solutions for the creation of depth maps and DIBR related to their framework of multi-view adaptation.

The proposition of [25] involves the use of forward warping in the construction of depth map. The following advantages have uniquely featured in this approach. First, re-sampling artifacts are filled in by inverse warping. Second, dis-occlusions are processed while omitting warping of edges at high discontinuities. Third, dis-occlusion in-painting approach explicitly uses depth information.

The discuss presented in this section will be taken into consideration in implementation to be found in section 4.4 in chapter four.

2.7 METRIC RECONSTRUCTION

One problem which has remained deeply rooted in multi-view geometry is the reconstruction of a scene starting with multiple 2D representation of different parts of the same scene. This is usually formulated as a non-linear problem which has to be solved using an iterative optimisation process starting from a sub-optimal solution obtained by using linear methods. In this chapter, the main focus is to perform Metric reconstruction of a 3D scene from multiple image samples using SBA algorithm. During this process the position, orientation and calibration of the cameras are recovered. SBA has a well-deserved reputation for almost always having been called upon to optimise both the 3D description of the scene geometry and camera viewing parameters in feature-based reconstruction which widely occur in theoretical and practical aspects of computer vision. In most of the multi-view cases which have been documented in literature, acquisition cameras are assumed to have different centres of projection. The focus here is on the unique problem of refining visual reconstruction which yields camera pose and calibration, and 3D structure estimate when multiple views have a single centres of projection. First an experiment in simulation is performed based on a synthetic scene. The second part is dedicated to real scenes.

2.7.1 *Introduction*

The objects of interest in the world are simply a set of points. The intensity acquisition of these points is achieved through the non-singular transformation capability of cameras which are almost usually based on signal processing principles, measurement, and algorithms [219]. Although important research and developmental effort have been targeted at camera fabrication based on state-of-the art technology, there still exist practical limits on the intrinsic parameters such as focal length, [220]. For example, in Figure 2.25, the dependence of two important parameters Depth of Field (DOF) and FOV on camera focal length f, working distance u, magnification μ, circle of confusion c, f-number of camera

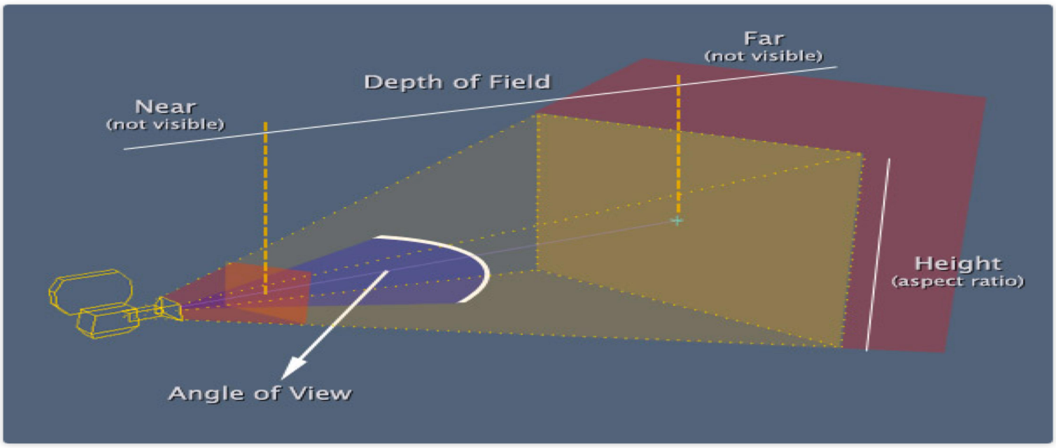


Figure 2.25: Dipiction of DOF and FOV. Taken from http://blog.db-in.com/cameras-on-opengl-es-2-x/.

lens f_n, and sensor physical size s are expressed in (2.20) and (2.21) respectively [220].

$$DOF = \frac{2f^2 f_n c\,(\mu + 1)}{f^2\mu^2 - f_n^2 c^2} \tag{2.20}$$

$$FOV = s \arctan\left(\frac{d\,(u - f)}{2uf}\right) \tag{2.21}$$

It is explicitly obvious from (2.20) and (2.21) that localization and matching errors [219], are deeply engraved in camera measurements. It turns out that the data acquired with such cameras are difficult to work with especially in scene reconstruction and other certain critical application areas, and often does not give accurate result. It is important that these errors be adequately attended to and considerably reduced in order to improve accuracy, reliability, and computational performance issues in image processing applications such as 3D reconstruction.

This work is intended to investigate the visual reconstruction of a scene from multiple images acquired using a camera with a SCOP as opposed to Multiple Centres of Projection (MCP) which is depicted in Figure 2.26 and Figure 2.27 respectively. Scene visual reconstruction by way of bundle adjustment attempts to recover a model of a 3D scene from multiple images [221]. As part of this, it usu-

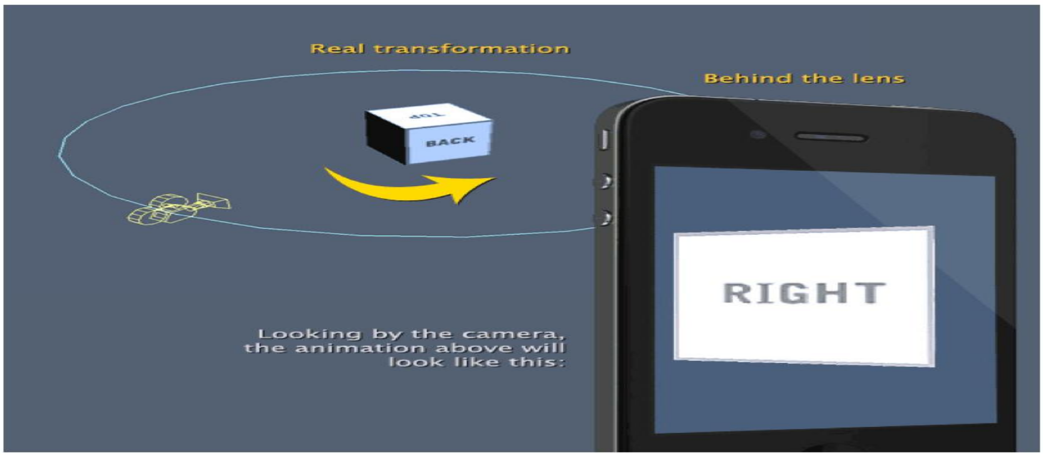


Figure 2.26: Single centre of projection. Taken from http://http://blog.db-in.com/cameras-on-opengl-es-2-x/.

ally also recovers the poses (positions and orientations) of the cameras that took the images, and information about their internal parameters [222–228]. Bundle adjustment can cope with any model that predicts the values of some known measurements or descriptors on the basis of some continuous parametric representation of the world, which is to be estimated from the measurements. It can incorporate several types of image observations along with their associated precisions and it can provide statistical information regarding the quality of the solution. Bundle adjustment tools can accommodate different sensor types, e.g. frame cameras and linear array scanners. Bundle adjustment methods can easily handle missing data, i.e. points which are not visible in a given image due to occlusion or insufficient object coverage. They can also handle redundant data, i.e. points which are visible in any number of images [229–236].

Since different parts of the scene are observed from the same point of different directions with a single camera, the determination of point correspondences becomes a major challenge. Also the camera parameters (intrinsic and extrinsic) and the 3D reconstructed points contribute to the computational complexity of the process [237]. The number of the camera parameters is directly decided by the number of images but the number of the 3D reconstructed points is largely decided by the image resolution and the scene texture. This problem is addressed by using optimization algorithm in the form of SBA to maximize the likelihood

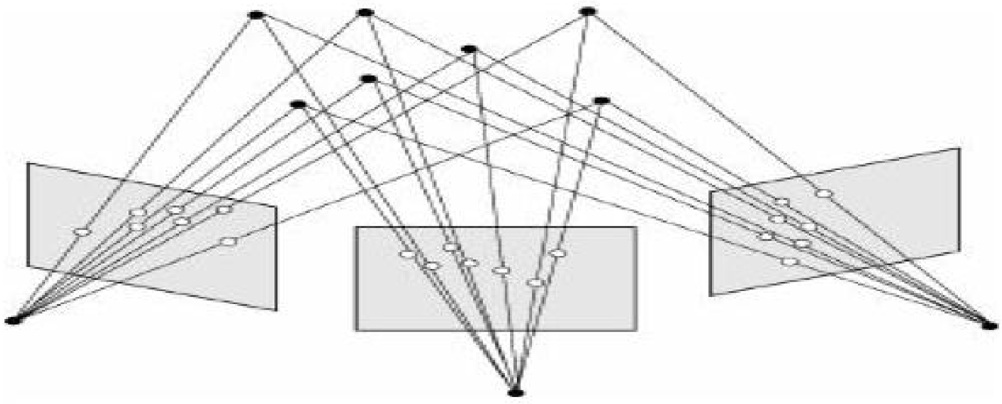


Figure 2.27: Illustration of multi-view with cameras with different centres of projection. Taken from [219]

of reconstruction. One interesting aspect of this study is that apart from intensity variation, there is a strong texture variation of the scene.

SBA is an active research area in computer vision. It provides a concise and straightforward introduction as well as a detailed coverage on the unique research problem of re-projection error minimization in a wide range of stereo imagery. It adequately addresses the problems of high computational and memory costs associated with least squares optimisation techniques used in the early years of bundle adjustment. This is in view of the fact that the Jacobian of some parameters exhibit sparsity which can be tailored to improve computational speed. This concept has been successfully applied in parameter refinement problems and has allowed inconsistencies between stereo pairs to be removed. It has also enhanced the realistic descriptions and modelling of other sophisticated imaging system applications [229, 237, 238].

There are two trends which have been widely adopted in scene reconstruction algorithms based on SBA. One situation is when the scene is fixed and the camera motion is around it. This is of course simple to implement with hand-held cameras, otherwise a specialised rig is required. When a hand–held camera is used, the recovery of homographies that maps images to each other and consequently allows the images to be transformed and combined becomes a huge challenge. Alternatively, the scene object can be placed on a turn-table and images are acquired at regular angular intervals with a fixed camera. This is only realisable if the scene is a portable object or cast object.

Understandably, [239], the stereo image data obtained by **HiRISE!** (**HiRISE!**) and used in the characterization of the surface, subsurface, and atmosphere of Mars have been optimized using SBA, in view of the scientific relevance of the project. The project is aimed at obtaining high-precision topographic information which is at the core of most space exploration programmes.

SBA is also known to have been used in image mosaicking [240]. This approach facilitates an accurate 3D reconstruction from multiple images of the same. This procedure features result which is statistically optimal based on the enforcement of hard geometric constraints. Frame decimation discussed in [241] provides for an automatic method which decides on the frame rate for any image sequence to be used in a structure from motion problem. This idea makes use of a coarse to fine, optical flow based video mosaicking algorithm.

Expansion procedure is adopted in [242] to output a qualitative dense set of patches on the surface of an object. Feature points obtained by matching multiple images are used to generate initial patches. These are continuously expanded until dense patches are obtained. The optimality of the patches is determined by local photometric consistency and global visibility constraints.

In Simultaneous Localization and Mapping (SLAM) problem, [243], classical Bundle Adjustment (BA) technique has been used to precisely describe and estimate the position of a mobile robot on a constructed map of the same environment.

Cost function is very critical in the formulation of SBA. The comprehensive theoretical background of SBA given in [244] defines a robust cost function as the square sum of all the dimensions of an error vector function. This is contrary to the conventional method of approximating the cost function locally with a quadratic Taylor expansion. The discussion also provides an entirely different proposition in which it is investigated that error build-up can be described as a function of accuracy that is obtainable through the use of bundle adjustment and helps to improve the reliability of camera tracking. The discussion also focuses on how bundle adjustment facilitates real-time application.

It is observed that the definition of variables in bundle adjustment is critical when large error propagation is needed in order to correct global error at loop closure. This will involve constraining the expected solution. An adaptive bundle adjustment is proposed in [245] which works in a metric-space defined by a connected Riemannian manifold. This is aimed at addressing the problem of single coordinate frame which is responsible for the high computational cost of BA.

Since line features from image correspondences is an integral part of scene modelling and augmented reality, [246], the 'two points' and the 'two plane' over parameterisation which can cause gauge freedoms and/or internal constraints is addressed with plucker coordinates so that the feature lines can be represented in 3D.

In [236], solution instability due to the linear dependencies between parameters when perspective algorithms (collinearity based) are used has been identified to be common in imaging situations where long focal length lenses and narrow FOV play a prominent role. This problem is addressed by the use of a scaled orthographic projection model based on linear algebraic formulations. Using quaternions, The mathematical model developed uses quaternions which translate to partial derivatives as well as the inner constraint equations for a scaled orthographic BA.

In the method adopted by [247], fast direct Cholesky decomposition techniques are employed to solve SBA problem with sparse linear sub-problem. The techniques involves the use of a compressed representation of large sparse matrices for efficiently handling the block data structures of SBA to take advantage of this representation. The performance evaluation is proved to surpass what is known to be obtainable from the method used in [219].

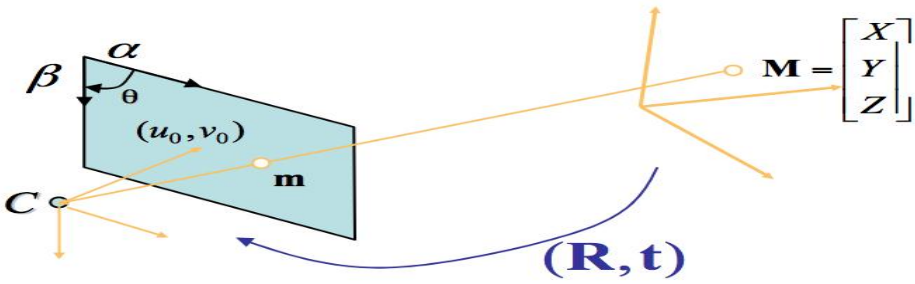


Figure 2.28: Perspective Camera Projection. Taken from [72].

2.7.2 *Bundle Adjustment Problem Formulation*

Bundle adjustment is a technique used to compute the maximum likelihood estimate of structure and motion from image feature correspondences. It exploits the sparse primary structure of the problem, where connections exist just between points and cameras. This is a non-linear system problem which have to be solved using an iterative optimization process starting from a sub-optimal solution obtained by using linear methods.

A camera can be modelled in several different ways. Affine and orthographic projections are sometimes useful for distant cameras, and more exotic models such as push-broom and rational polynomial cameras are needed for certain applications [229]. Other camera models can be derived from it. In addition to pose (position and orientation), and simple internal parameters such as focal length and principal point, real cameras also require various types of additional parameters to model internal aberrations such as radial distortion. However, perspective projection as shown in Figure 2.28, [72], is the most prominent. Perspective projection is the linear mapping between the extended coordinates of any world point M and its corresponding image point m such that collinearity property exists between M, m, and C (centre of projection) . This can be expressed as

$$\Lambda m = QM = P \tag{2.22}$$

where M is 4×1 vector and m is 3×1 vector. Λ is an arbitrary scale factor. Q is a 3×4 vector referred to as the projection matrix. Therefore, in reality, an object or structure consists of several M points and image of such an object or structure will consist of corresponding number of m points. An important characteristic which can be exhibited by a perspective camera is for the first three left columns of M to be non-singular. It therefore means Q can be further decomposed such that

$$\Lambda m = K\,[R\,|\,t]\,M \tag{2.23}$$

K is 3×3 upper triangular matrix. It is called camera or calibration matrix of the camera. It comprises of the optical properties of the camera namely, focal length, principal point and aspect ratio. R is an orthogonal 3×3 matrix and t a 3×1 vector. R and t are collectively referred to as the camera's extrinsic orientation and correspond, respectively, to the rotation and translation that make up the rigid transformation from the world to the camera coordinate frame [219]. The coordinate system C attached to the camera is related to the world coordinate system through a rotation R followed by a translation t.

For a multi-view setting, consider having M_j scene points captured by several cameras described by Q^i. Assuming the projection of M_j point due to Q^i camera is m^i_j. Starting from multi-view image samples, multi-view 3D reconstruction involves the determination of M_j and Q^i such that (2.22) is satisfied as expressed in (2.24).

$$\Lambda m^i_j = Q^i M_j \tag{2.24}$$

A significant challenge in SBA is that (2.24) is not exactly satisfied. m^i_j has inherent noise superimposed during the measurement process. Therefore, for every image point m^i_j, a predictive model $\hat{m}^i_j = m\left(Q^i, M_j\right)$, [229], is required such that

$$\Delta\hat{m}^i_j \equiv m^i_j - m\left(Q^i, M_j\right) \tag{2.25}$$

$$f\left(Q, M\right) = \sum_{i=1}^{c}\sum_{j=1}^{d} V_{ij}\left\|m^i_j - m\left(Q^i, M_j\right)\right\|^2 \tag{2.26}$$

Equation (2.25) is referred to as re-projection error. Hence the problem of scene reconstruction and camera parameter estimation boils down to the minimisation of the re-projection error between the image locations of observed and predicted image points, which is expressed as the sum of squares of a large number of non-linear, real-valued functions. The objective function for the minimization problem defined in the context of bundle adjustment is expressed in (2.26). Where c is the number of scene points and d is the number of cameras. V_{ij} is a visibility weight which equals 1 if a scene point j can be seen in camera i, otherwise it equals 0. If the unexpected variation in pixel coordinates is modelled as Gaussian noise with zero mean, then (2.26) becomes a statistical non-linear model. Using the condition of linear independence of the columns of Q^i, (2.26) can be expressed as a set of linear equations as in (2.27)

$$\left(Q^i\right)^T Q^i \hat{f} = \left(Q^i\right)^T m_j^i \tag{2.27}$$

whose parameters can be estimated using least-squares algorithm like Levenberg-Marquardt Algorithm (LMA). However, for a large set of object points and camera parameters which, constitute the unknown contributing to the minimized re-projection error, the system represented by (2.27) becomes overdetermined. The computational cost will then have cubic complexity [245]. Therefore, a specialised LMA known as SBA is required in order to seek a minimal solution.

2.7.3 *Formulation of SBA Based on Levenberg-Marquardt Algorithm*

Problems arising in the contest of SBA have the possibility of being solved using conjugate gradient technique. However, LMA remains widely known and attractive at solving unconstrained non-linear least squares problems [248]. It can take advantage of the sparseness that exists at different levels in bundle adjustment. LMA can be thought of as a combination of steepest descent and the Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: Slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a

Newton's method. The behaviour, in terms of convergence, and computational cost of this process depend on the parameterisation used to represent the problem, i.e of structure and motion [219].

Starting from the measured image pixel coordinates m resulting from image formation by multi-view cameras with parameter $P\epsilon\mathbb{R}$, an estimated measurement vector $\hat{m}\epsilon\mathbb{R}$ can be formulated as

$$\hat{m} = g\left(P\right) \tag{2.28}$$

Let the difference between the measured quantity m and estimated quantity $\hat{m}$ be

$$m - \hat{m} = \varepsilon \tag{2.29}$$

A minima value for the quantity $\varepsilon^T\varepsilon$ known as square distance can be used as a criteria in the determination of P^+ which satisfies the relation of (2.28) if an initial guess value P_0 is provided.

The fundamental basis of LMA is that for an infinitesimal change δ_P in P, g can be approximately expressed as

$$g\left(P + \delta_P\right) = g\left(P\right) + J\delta_P \tag{2.30}$$

where J is the matrix of all first-order partial derivatives of the function g. It is called Jacobian. Equation (2.29) can now be expressed as

$$\|m - g\left(P + \delta_P\right)\| = \|\varepsilon - J\delta_P\| \tag{2.31}$$

For each δ_P, an iteration process is performed using (2.31) until when $J\delta_P - \varepsilon$ is orthogonal to the column space of J, [228]. This means that

$$J^T\left(J\delta_P - \varepsilon\right) = J^T J\delta_P - J^T\varepsilon = 0 \tag{2.32}$$

$$J^T J\delta_P = J^T\varepsilon \tag{2.33}$$

J^T is the transpose of J. Equation (2.32) is referred as normal equations, [219, 249]. J^TJ constitute the Hessian of $\frac{1}{2}\varepsilon^T\varepsilon$ and the infinitesimal change δ_P is called Gauss-Newton step. $J^T\varepsilon$ is the steepest descent components.

For the purpose of error reduction and to make sure that J is not rank deficient or singularity of J^TJ, a modified form of (2.33) referred as the augmented normal equations is solved as shown in (2.34). This involves the manipulation of the diagonal elements of J^TJ using σ as a damping element. σ is always chosen such that $J^TJ + \sigma I$ is non-singular and positive definite.

$$\left(J^TJ + \sigma I\right)\delta_P = J^T\varepsilon \tag{2.34}$$

The unpredictability of m described by covariance matrix $\sum_m$, can be factored into (2.34) such that we now have $\left(J^T \sum_m^{-1} J + \sigma I\right)\delta_P = J^T \sum_m^{-1} \varepsilon$. This variation allows for minimisation of squared $\sum_m^{-1} -norm$ (Mahalanobis distance) as

$$\|\varepsilon\|^2_{\sum_m} = \varepsilon^T \sum_m^{-1} \varepsilon \tag{2.35}$$

The computational complexity is still cubic.

In multi-view setting, the inherent characteristic of lack of interaction among parameters for different 3D points and cameras results in the underlying normal equations exhibiting a sparse block structure [219]. Therefore, SBA is a variant of LMA which seeks to reduce the computational cost of bundle adjustment by taking advantage of data sparseness. It is established in literature that SBA can cope with parameterisation of the multi-view reconstruction problem such as arbitrary projective cameras, partially or fully intrinsically calibrated cameras, exterior orientation (i.e. pose) estimation from fixed 3D points, and refinement of intrinsic parameters.

By combining the simplified forms of the right and left hand sides of $\left[J^T \sum_m^{-1} J + \sigma I\right] \begin{bmatrix} \delta_a \\ \delta_b \end{bmatrix} = J^T \sum_m^{-1} \varepsilon$, we have

$$\begin{bmatrix} X_1 & 0 & . & 0 & Z_1^1 & Z_2^1 & . & Z_c^1 \\ 0 & X_2 & . & 0 & Z_1^2 & Z_2^2 & . & Z_c^2 \\ . & . & . & 0 & . & . & . & . \\ 0 & 0 & 0 & X_d & Z_1^d & Z_2^d & . & Z_c^d \\ \left(Z_1^1\right)^T & \left(Z_1^2\right)^T & . & \left(Z_1^d\right)^T & Y_1 & 0 & . & 0 \\ \left(Z_2^1\right)^T & \left(Z_2^2\right)^T & . & \left(Z_2^d\right)^T & 0 & Y_2 & . & 0 \\ . & . & . & . & . & . & . & 0 \\ \left(Z_c^1\right)^T & \left(Z_c^2\right)^T & . & \left(Z_c^d\right)^T & 0 & 0 & 0 & Y_c \end{bmatrix} \begin{bmatrix} \delta_{a_1} \\ . \\ . \\ \delta_{a_d} \\ \delta_{b_1} \\ . \\ . \\ \delta_{b_c} \end{bmatrix} = \begin{bmatrix} \varepsilon_{a_1} \\ . \\ . \\ \varepsilon_{a_d} \\ \varepsilon_{b_1} \\ . \\ . \\ \varepsilon_{b_c} \end{bmatrix} \qquad (2.36)$$

Equation (2.36) is the partitioned form of the normal equations. We can simply write (2.36) as

$$DE = F \qquad (2.37)$$

Based on orthogonality principle, the identity matrix in the upper left and lower right corners of D can be written as A^TA and B^TB respectively. The upper right corner containing Z elements equals A^TB while the lower left corner with Z elements equals B^TA, i.e the transpose of A^TB. For the purpose of further simplification, (2.36) is written in compact form as

$$\begin{bmatrix} X^* & Z \\ Z^T & Y^* \end{bmatrix} \begin{bmatrix} \Delta_a \\ \Delta_b \end{bmatrix} = \begin{bmatrix} \varepsilon_a \\ \varepsilon_b \end{bmatrix} \qquad (2.38)$$

where

$$X = \sum_i^d A_i^T A_i \qquad (2.39)$$

$$Y = diagonal\left(B_1^T B_1, B_2^T B_2, ..., B_{c-1}^T B_{c-1}, B_c^T B_c\right) \qquad (2.40)$$

$$Z = \left[A_1^T B_1, A_2^T B_2, ...A_c^T B_c\right] \qquad (2.41)$$

$$\varepsilon_a = \sum_i^d A_i^T \varepsilon_{a_i} \tag{2.42}$$

$$\varepsilon_b = \left[B_1^T \varepsilon_{b_1}, B_2^T \varepsilon_{b_2}, \dots . B_{c-1}^T \varepsilon_{b_{c-1}}, B_c^T \varepsilon_{b_c} \right]^T \tag{2.43}$$

$\Delta_a = \delta_a$ and $\Delta_b = \delta_b$. X^* and Y^* are the augmented forms of X and Y respectively. The pre-multiplication of (2.38) by

$$\begin{bmatrix} I & -ZY^{*(-1)} \\ 0 & I \end{bmatrix} \tag{2.44}$$

gives

$$\begin{bmatrix} \left(X^* - ZY^{*(-1)} \right) & 0 \\ Z^T & Y^* \end{bmatrix} \begin{bmatrix} \Delta_a \\ \Delta_b \end{bmatrix} = \begin{bmatrix} \varepsilon_a - \left(ZY^{*(-1)} \varepsilon_b \right) \\ \varepsilon_b \end{bmatrix} \tag{2.45}$$

The simplification of (2.45) gives two simultaneous equations. If we make Δ_a the subject of the formula, we get

$$\Delta_a = \frac{\varepsilon_a - \left(ZY^{*(-1)} \varepsilon_b \right)}{\left(X^* - ZY^{*(-1)} Z^T \right)} \tag{2.46}$$

For Δ_b we have

$$\Delta_b = \frac{\varepsilon_b - \Delta_a Z^T}{Y^*} \tag{2.47}$$

An important observation is that the invertible form of Y^* can be seamlessly performed since it is block diagonal. Δ_a and Δ_b can now be used to update P such that

$$P_{update} = P + \begin{bmatrix} \Delta_a^T & \Delta_b^T \end{bmatrix}^T \tag{2.48}$$

P_{update} is further used to update the re-projection error ε expressed in (2.29), i.e

$$\varepsilon_{update} = m - g\left(P_{update} \right) \tag{2.49}$$

Therefore, due to the sparseness of the primary structure, the implementation steps of BA are entirely linearised. Hence a computational complexity of $\mathcal{O}(c)$ is exhibited [219].

We now test ε_{update} by comparing its absolute value with the absolute value of ε. If ε_{update} is less or equal to ε, i.e $\left(\left\|\varepsilon_{update}\right\| \leq \|\varepsilon\|\right)$, σ is updated, otherwise the iteration process continues until convergence is achieved.

Chapter five is based on the theoretical concept discussed in this section.

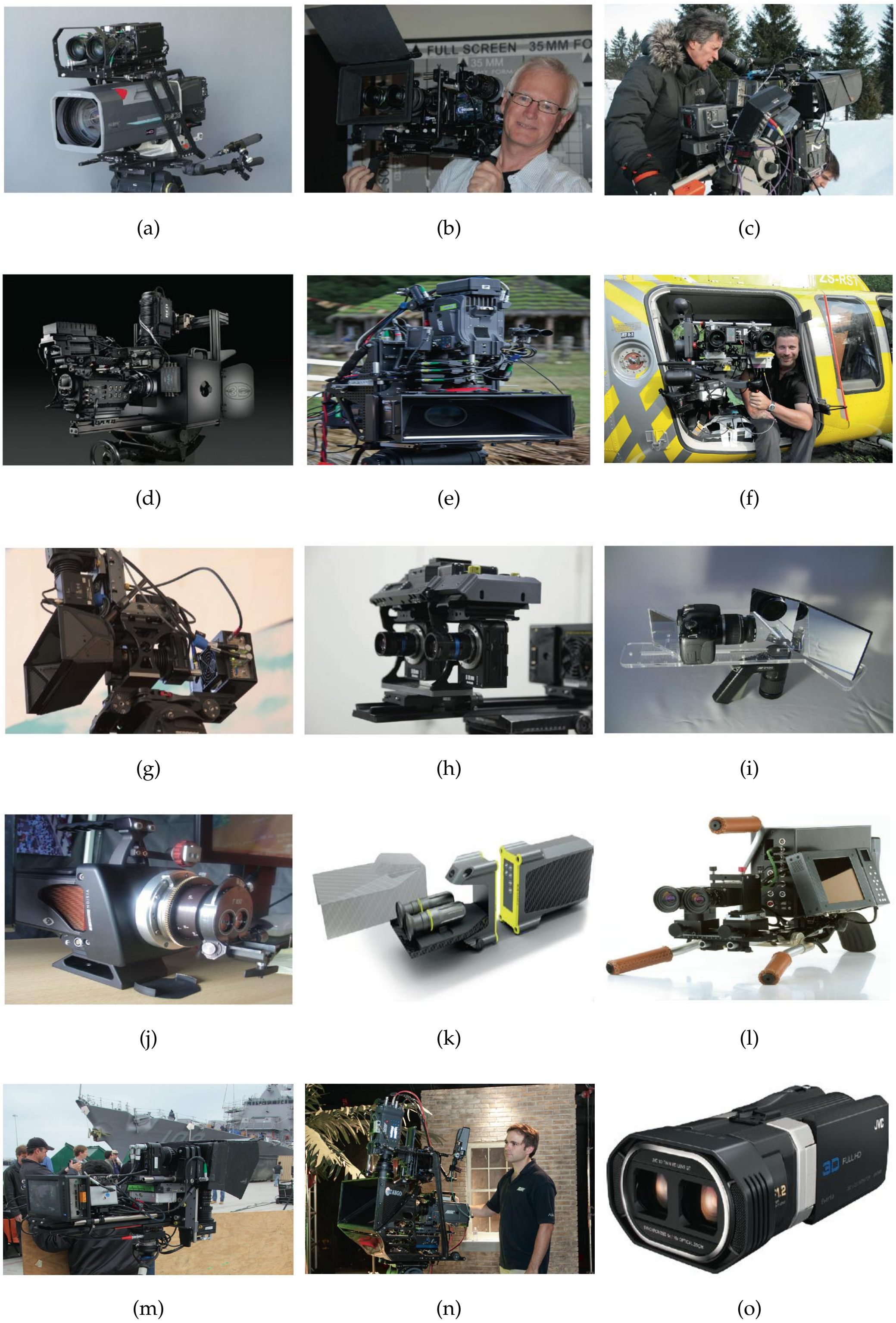
2.8 MULTI-CAMERA CONFIGURATION

3D content, [250], can be obtained by having a high-resolution, wide-angle camera focused during a moderate object motion [73]. However, in computer vision, a synchronised set of multi-cameras with known accurate positions and orientations, brightness and chromatic characteristics are used to observe object surface areas. Single camera techniques, holoscopic capture devices, pattern projection technique, and time-of-flight techniques have been actively used in 3D content acquisition in other applications.

The work by [251] in which it is demonstrated that combination of a lenticular array with photographic film in order to capture and reproduce stereoscopic imaging, is known to be the beginning of adopting multiple cameras to capture a scene. The choice of camera types [42] as shown in Figure 2.29, and configuration for multi-view video has become a hot topic in computer vision in recent times. It has many commercial and military applications, from video surveillance to smart home systems, from traffic monitoring to oil exploration.

Complexity of the scene, self-occlusion and mutual occlusion of moving objects, diverse sensor properties are prominent factors that can affect the performance of any chosen configurations. Traditional configurations can be classified into different categories according to the shape traced out by the cameras. Linear array is the simplest configuration.

One important application of multi-camera system is in multi-view video as depicted in Figure 2.30. This is popularly used in the creation of 3Deffect, [24, 252], which enables a 3D scene to be viewed by freely changing our view-



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

Figure 2.29: Different camera types, rigs, and prototypes. Taken from [42].

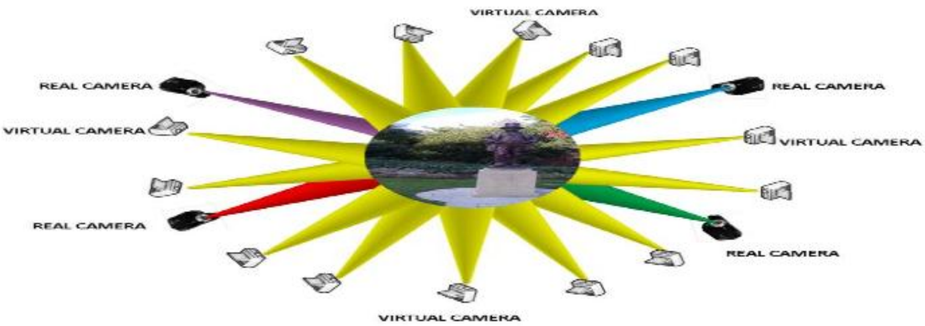


Figure 2.30: Figure 2.29: Multi-camera technique used in multi-view video.

point, and 3DTV in which the illusion of depth is created. Multi-view imaging has also attracted increasing attention in another wide variety of interesting new research topics and applications. These range from virtual view synthesis, high performance imaging, image and video segmentation, object tracking and recognition, environmental surveillance, remote education, to industrial inspection [2]. In video content service such as video summarisation, [253], a condensed form of video content is generated for the purpose of browsing, retrieval and storage enhancement. 3D seismology now has a considerable driving force as opposed to its two-dimensional counterpart. It can help to solve the increased dimensionality of the problems associated with imaging, processing, and visualization of resultant images.

Until recently, however, the focus has been on understanding the successes of three lines of development in camera configurations such as parallel array, convergence array, and divergence array, Figure 2.31. Parallel array is the simplest form in which identical cameras are all in a linear orientation [254]. It is mentioned in [255] that more complicated settings can have different camera lens properties and zoom facing the same 3D scene from different directions. The geometry of these camera topologies can very easily be analysed. The need to improve the camera architecture used in visual acquisition has been made evident from a variety of angles. This includes the issue of camera density, the reduction in the number of physical cameras, image quality, synchronization [11], depth estimation and occlusion [256, 257].

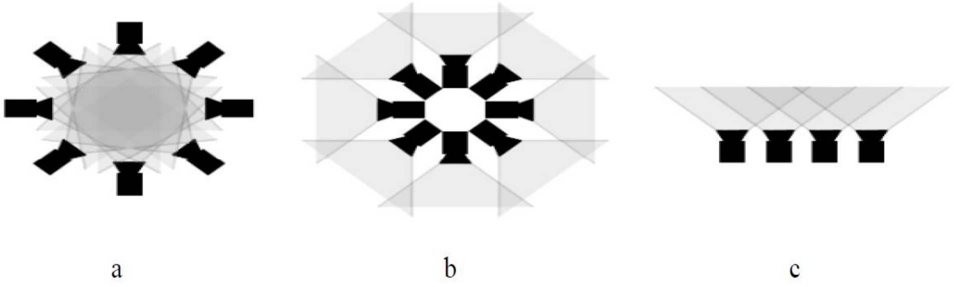


Figure 2.31: Multi-view camera arrangements: (a) converging, (b) diverging, and (c) parallel arrangement. Taken from [24].

Fairly recently, in [220], other issues such as visual attention has been considered as an important aspect of perception and its understanding is therefore necessary in the creation of 3D content. Reference [258] has targeted an important discussion at remapping the disparity range of stereoscopic images and video. This is aimed at reducing the effect of a complex combination of perceptual, technological, and artistic constraints on the displayed depth and the resulting 3D viewing experience. A promising approach to measure 3D visual fatigue using electroencephalogram and event-related potential has been proposed by [259].

The challenge of camera placement in multi-camera setting has been highlighted in [260]. In the work presented in this thesis, the description of a multi-camera topology: the TCA, for acquisition of visual content is proposed. This is in spite of the fact that best observability of the object surface with a single ring camera arrangement can be achieved when the ring is at mid-height of the target object [73]. The strong point of TCA is that it is based on an efficient trapezoid which is half of a regular hexagon. More importantly, a trapezoid defines four of the six sides of a frustum (hexahedral) which has become an interesting topic of intense research in mesh generation. In [44], the algorithm employs geometrical, optical, and reconstruction constraints to realise complete scene coverage with minimum number of cameras. A trapezoid is a quadrilateral with a pair (or at least one) of opposite parallel sides [261]. Acquisition cameras can be arranged on the sides of a trapezoid as shown in Figure 2.32. Though both conceptual

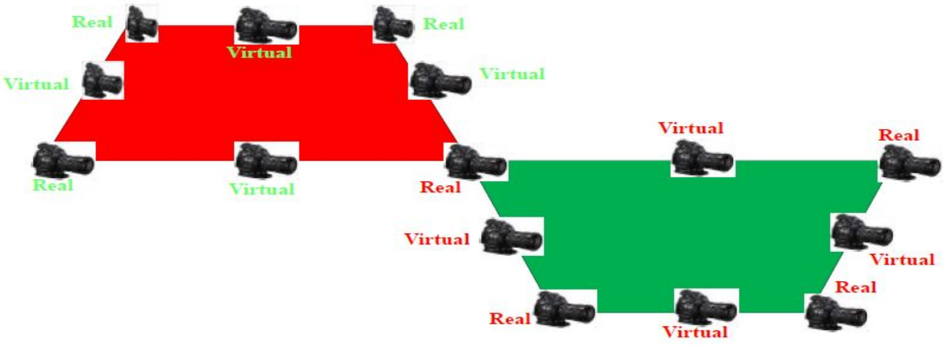


Figure 2.32: Trapezoidal camera architecture.

and implementation challenges are in view, the architecture of Figure 2.32 can be implemented at a certain frequency around a scene as depicted in Figure 6.1.

The proposition in [262], is made of a scalable architecture for a distributed image-based rendering system based on a large number of densely spaced video cameras. This way, the task of dynamic geometry creation is eliminated and allows the application of light field rendering techniques. The used light field rendering algorithm helped to reduce bandwidth issues and to provide a scalable system.

In the self-reconfigurable camera array discussed in [263], 48 cameras are involved and mounted on mobile platform. These cameras captured images at a resolution of 320×240. A 100 Mbps Ethernet connection allowed 48 cameras to send image sequences to the computer simultaneously at $15 - 20$ fps. The proposed real-time rendering algorithm is implemented in software. This proposition is characterised by a novel self-reconfiguration algorithm to move the cameras, and achieve better rendering quality than static camera arrays.

An array of plenoptic cameras is used to capture light fields as described in [264] many other similar work. A main lens and a set of lenslets are used. Beyond the focal point of the main lens is where the lenslet array is positioned. This approach ensures that non-overlapping content interleaved in the pixels under the lenslets can be captured. Also fusion super-resolution can be realised. The image acquired by the entire system is located at the focal point of the lenslet array. Whist this system can be said to be a build up of the first idea of plenoptic camera, its resolution suffers a serious reduction because summation of the shifted versions of the sub-aperture images is performed in order to achieve refocusing ability. Fundamentally, different sub-aperture of the main lens is captured by different lenslet and provides a depth of field that is correspondingly larger than that produced by the full aperture of the lens.

In another similar work by [265], the emphasis is to extend the depth of field by shrinking the aperture and the focal length of each lens in the array. Furthermore, advantage is taken of the considerable overlapping fields of view exhibited by the lenses, the resolution encoded in the downsampled and aliased images of the array can be recovered.

The subdivison of the main camera aperture into smaller independent channels with almost the same optical property is the approach at the centre of the work presented in [266]. The resultant system is a compact image-capturing system consisting of Thin Observation Module by Bound Optics (TOMBO). The TOMBO is basically a combination of a multiple-imaging system (compound-eye imaging system) and post-digital processing, can provide a compact hardware configuration as well as processing flexibility. It aims to reduce the track length of an electronic imaging system. An image of higher resolution compared to any of the multiple sets of elemental optics each of which consists of a microlens and a photosensitive cell, is obtained using Iterative Back-Reprojection (IBP) approach.

With the simple and robust method proposed in [121], it is possible to generate new views from virtual viewpoints without depth information or feature matching. This is can be done simply by combining and re-sampling the available images. The input images obtained from camera arrays are considered as 2D slices of a 4D function.

In [267] an irregular lens arrangement was used to ensure that component images were non-identical and used super-resolution algorithms to restore a high-resolution image. However, neither of these approaches considered an integrated approach to parallax detection/correction and super-resolution, choosing instead to recover the resolution at a fixed depth from the camera. In addition, the demonstrated resolutions were low compared to today's expectations for a mobile imaging system.

2.8.1 *Conventional Camera Architecture*

A brief review of the known implemented camera architectures namely: parallel, convergence, and divergence, is presented in this section.

2.8.1.1 *Parallel Array*

According to [268], parallel arrangement of cameras allows for wide angle capturing of the scene. It is also known to feature simple disparity calculation. A hybrid camera system consisting of five high-definition video cameras arranged in a linear array and one time-of-flight depth camera for the generation of multi-view video has been proposed by [269]. The merit of this technique is that the initial depth map at each viewpoint, obtained through 3D warping operation, is further optimized using segment-based stereo matching.

Camera Array Pursuits for Plenoptic Acquisition (CAPPA), [217], have been constructed using Sony XC-333 cameras. It is aimed at capturing multi-view video. Dense and sparse camera arrangement is realized using a modular unit designed for this purpose. Video from 16 cameras is translated to four-screen sequences using four SonyYS-Q430 (quad microprocessor). A final single 16-screen sequence is generated using a fifth quad processor to combine the earlier four-screen sequences.

Reference [270] has used an array of 64 cameras, Intel Xeon 5160 dual-processor machine, and an NVIDIA GeForce 8800 Ultra graphics card to capture multi-view video in real time. It is also characterized by an interactive control of viewing parameters.

Sunex DSL841B lenses with a focal length of 6.1mm and Marshall Electronics V-4350-2.5 lenses were used in a linear array of cameras to capture indoor and outdoor scenes respectively. In two separate experiments, 128 and 48 camera systems were used. Considerable implementation performance and improved image quality have been demonstrated through this experimental setup.

In the work for which [271] is famous, video light fields, high-dynamic-range video, high-resolution panoramas, and ultra-high speed video were generated. In the particular cases of high-dynamic-range video, high-resolution panoramas, and ultra-high speed video, varying of exposure times, splaying of direction of view, and staggering of camera triggering times were respectively performed.

2.8.1.2 *Convergence Array*

It is interestingly observed that the convergence camera arrangement provides detailed information about a scene or an object. Convergence camera configuration has been used in the experimental system of FTV, [272] in order to acquire high-resolution video and analogue signal up to 96 kHz. It is a "100-camera system" JAI PULNiX TM-1400CL developed at Nagoya University and Tanimoto Laboratory.

In another experiment, [273], a stadium is surrounded with eight texture acquisition video cameras (SONY DXC-9000) which are capable of performing a progressive scan. A similar camera positioned on the stadium ceiling acquires the Z component of a player's position. In this experimentation, the FOV of the horizontal plane cameras are controlled by the FOV of the single vertical plane camera. This is done to extend the FOV of the multiple cameras to the stadium areas that were not initially covered.

2.8.1.3 *Divergence Array*

The simplest divergent camera configuration consists of a camera usually panned horizontally or vertically at certain intervals through an angle of 360 degree [274]. Each image sample contains both the geometric and optical properties of different parts of the scene. A panoramic view of the scene is finally constructed by the method of image stitching.

In concentric mosaicking, images are acquired using cameras that are equally spaced out on the circumference of concentric circles. Therefore, all the cameras not only have the same centre of projection, [198, 205] but all input image rays are

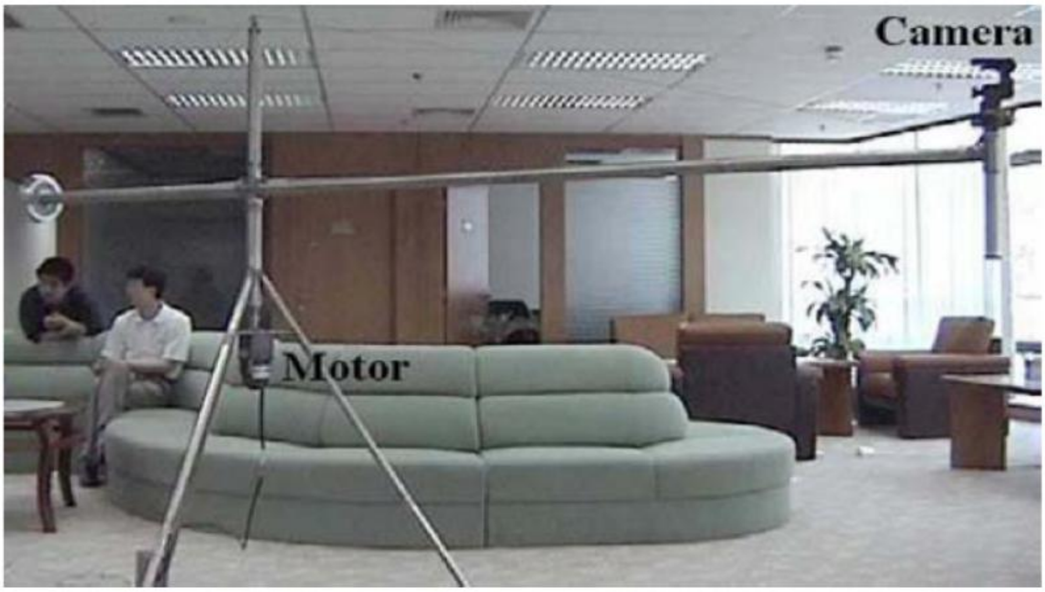


Figure 2.33: Concentric mosaicking. Taken from [112].

naturally indexed in radius, rotation angle, and vertical elevation. An example is shown in Figure 2.33.

For the multi-view video recording in [275], IEEE1394 cameras are placed in a convergent setup around the centre of the scene. The video sequences used are recorded from eight static viewing positions arranged at approximately equal angles and distances around the centre of the room. The cameras are synchronised via an external trigger and all the video data is directly streamed to the hard drives of four control PCs, each of which is connected to two cameras. Video frames are recorded at a resolution of 320×240 pixels. The frame rate is fundamentally limited to 15 fps by the external trigger. The cameras' intrinsic and extrinsic parameters are determined, thereby calibrating every camera into a common global coordinate system.

The discussion in this section is applied in chapter six.

2.9 CRITICAL ANALYSIS

Since the environment and infrastructure at our disposal are favourable for computing and information processing more than ever before, it stands to reason that the rendering or reconstruction of a scene from multiple image sample of the same scene is a possibility but could be challenging. In [165], It is emphasised that new views of scenes are be created from a collection of densely sampled images or videos. This understanding is also shared in [24], however, concerns were further raised about the synchronisation of the cameras. Synchronisation requirement means that the multiple cameras be triggered simultaneously most especially when the scene is dynamic. Other important requirements such as the geometric and photometric calibration of the cameras have also been mentioned to be critical. There could be a considerable error propagation in the sequential process leading to the recovery of camera intrinsic parameters and image rendering. Another important requirement highlighted in [24] which almost always facilitate the reconstruction of a 3D scene with absolute depth is the visibility of object surface area in at least two cameras.

To reasonably address the issues of synchronisation and camera calibrations as a research gap, the use of virtual cameras is proposed which can provide for a reduction in the number of physical cameras used. Hence, this is a cost effective alternative which can offer a reasonable degree of accuracy. A virtual camera can be realised using the method of IBR which has been extensively discussed in the earlier part of this chapter. The categories of IBR with and without geometry and the combination of the two are to be used.

Another research gap is that images with multiple centre of projection are have been widely used in 3D reconstruction. As part of the work presented in this thesis, a 3D reconstruction of a scene based on the use of multiple 2D images with a single centre of projection is proposed. This is in view of the fact that images usually used in the construction of panoramic images are often captured with a single camera mounted on a tripod.

Further more, the issue of visibility is challenged by the introduction of a trapezoidal camera architecture. In this configuration, few number of physical cameras are arranged at the edge of a trapezoid. The trapezoid graph allows for many ways of determining the position of a camera. Also, a point in the scene can be viewed from several viewpoint on the edge of the trapezoid graph.

2.10 CHAPTER SUMMARY

As part of the literature review discussion provided in this chapter, the concept of IBR has been highlighted. The criteria of geometry has been used for its categorisation. It comprises of IBR with explicit geometry, IBR with implicit geometry, and IBR with no geometry. IBR with explicit and no geometry are the two categorisations to be used in this work. A tailored discussion of IBR with regards to the original contributions of this thesis has also been given. Such contributions include, metric aspect of IBR, 3D panoramic image, computation of virtual environment, 3D scene reconstruction from image samples, and multi-camera configuration called trapezoidal camera architecture. Under the critical analysis section, a discussion leading to the definition of the research gaps and the suitable methods which could address the issues are highlighted.

3 IMAGE FORMATION

3.1 INTRODUCTION

In general, analysing cameras is a difficult problem and solutions are often found only for geometric approach. Geometric modelling goes beyond the pinhole perspective model. There exist many other types of simple camera models that are often used for modelling various imaging systems under different practical conditions. In this chapter, the image capturing capability of a camera is presented from optical perspective. Since most compact cameras can acquire only visible light, the description and propagation method of the visible part of the Electromagnetic (EM) spectrum reflected by a scene object is made starting from Maxwell's equations. We then seek to use this understanding in the modelling of the image formation process of the camera. The perception of the modelling presented here is that the type and size of the camera aperture determines the quality of the image formed. Also, the wave optics model leads to an important camera and image quality parameter called MTF. The model presented is based on a wave optics in which the wavefront is modified by the lens after diffraction has taken place at the camera aperture positioned at the front focal point of the lens.

3.2 BACKGROUND

The study of optics is believed to have started with the Sumerian's about 4000 B.C. The discovery of rock crystal lens, palaeolithic wall paintings often found in caves of almost total darkness, and the use of a lamp and a hand-held mirrors were significant developments which point to the cultivated high level civilisation during that time [186]. The idea about vision is arguably considered to have began in Greece by Euclid and others. Their understanding of vision was that humans see because some rays called eye rays emanate from the eye, strike the scene objects and are returned to the eye.

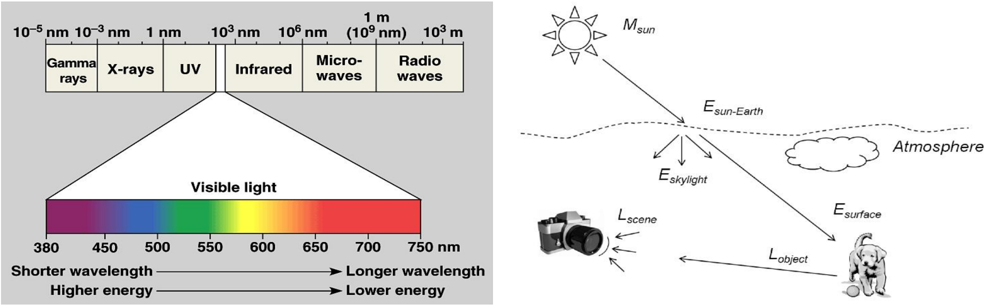
The second century A.D. witnessed a major conceptual shift when more quantitative experiments started. Refraction of light at the air-water interface was studied by Claudius Ptolemy [186].

Conventional imaging, and especially computer-assisted imaging, has brought many areas of research and applications to a mature plateau with astonishing advantages [276]. Therefore, we argue that camera modelling should be primarily directed towards describing mathematically any improvement strategies that can be made to the image formation process, and noise characterisation and evaluation [277, 278].

In a compact (refined) camera, the numerical form of the continuous variation of the scene object being sensed is produced. Comprehensively speaking, the modelling pipeline of the image formation process of a camera consists of radiometry, the camera (optics and sensor), the motion associated with the camera, the processing, the display, and the interpretation as functional blocks [278]. Only the optics and sensor model will be presented based on wave optics and the use of rectangular aperture.

3.3 IMAGE FORMATION BASED ON WAVE OPTICS

The energy that is captured by the camera is the visible part of the EM spectrum, a self-propagating wave comprised of oscillating electric and magnetic fields generated by the acceleration of charged particles [279]. The propagation of EM radiation through inhomogeneous media with random fluctuations of local optical characteristics results in the formation of an optical wave that is characterised by random temporal and spatial distributions of its parameters such as intensity, phase, and, in general cases, its state of polarisation. Theoretical modelling of the propagation and the diffraction of electromagnetic wave at the camera aperture provides a vital input into the understanding of established imaging techniques and the development of new procedures.



(a) (b)

Figure 3.1: (a) Visible light as part of electromagnetic spectrum. Taken from bio1151.nicerweb.com (b) the sun as a radiant source. Taken from [279].

Since the visible light shown in Figure 3.1 (a), reflected from scene objects, propagated and captured by the camera as depicted in Figure 3.1 (b), [279], is an EM phenomenon of a particular wavelength, it can be described by Maxwell's equations. Maxwell's equations represent a unification of the works of Lorentz, Faraday, Ampere, and Gauss that predict the propagation of EM waves in free space at the speed of light [280–283]. Maxwell's equations which model EM waves are stated as:

$$\nabla . \overrightarrow{E} = \frac{\rho}{\varepsilon} \quad Gauss's\, law \tag{3.1}$$

$$\nabla \times \overrightarrow{E} = -\frac{\partial \overrightarrow{B}}{\partial t} \qquad Faraday's\, law \tag{3.2}$$

$$\nabla \times \overrightarrow{H} = \frac{\partial \overrightarrow{D}}{\partial t} + \overrightarrow{J} \qquad Ampere's\, law \tag{3.3}$$

$$\nabla . \overrightarrow{B} = 0 \qquad Flux\, law \tag{3.4}$$

where∇ is the *del* operator, $\overrightarrow{E}$ and $\overrightarrow{B}$ are electric and magnetic field respectively. $\overrightarrow{D}$ is the electric displacement, $\overrightarrow{H}$ defines the magnetic intensity, $\overrightarrow{J}$ is the conduction current density in a volume, ε is the permittivity, and ρ is the charge density.

Equations (3.2) and (3.4) are valid only if position and time dependent magnetic $\overrightarrow{A}\left(\overrightarrow{r},t\right)$ and electric $\varphi\left(\overrightarrow{r},t\right)$ potentials exist at a field point on the camera sensor such that:

$$\overrightarrow{B} = \nabla \times \overrightarrow{A} \tag{3.5}$$

$$\overrightarrow{E} = -\nabla\varphi - \frac{\partial \overrightarrow{A}}{\partial t}. \tag{3.6}$$

Substituting $\overrightarrow{E}$ from (3.6) in (3.1), we get

$$\nabla.\left(-\nabla\varphi - \frac{\partial \overrightarrow{A}}{\partial t}\right) = \frac{\rho}{\varepsilon} \tag{3.7}$$

$$\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi = \frac{\rho}{\varepsilon}. \tag{3.8}$$

Using (3.5) in (3.2) we obtain

$$\nabla \times \overrightarrow{E} = -\frac{\partial\left(\nabla \times \overrightarrow{A}\right)}{\partial t} = -\nabla \times \frac{\partial \overrightarrow{A}}{\partial t}. \tag{3.9}$$

Rearranging (3.9) gives

$$\nabla \times \left(\overrightarrow{E} + \frac{\partial \overrightarrow{A}}{\partial t}\right) = 0. \tag{3.10}$$

Two constitutive relationships can written. These are

$$\overrightarrow{D} = \varepsilon \overrightarrow{E} \tag{3.11}$$

$$\overrightarrow{B} = \mu \overrightarrow{H}. \tag{3.12}$$

Equation (3.10) represents the infinitesimal rotation of the electric field. Since it is irrotational, it means it can be expressed as the gradient of electric potential $\varphi\left(\overrightarrow{r},t\right)$. Hence we can write

$$\overrightarrow{E} + \frac{\partial \overrightarrow{A}}{\partial t} = -\nabla\varphi. \tag{3.13}$$

Since the magnetic and electric potentials are not uniquely defined, there is the need to impose a constraint based on the gauge invariance of Maxwell's equations. The constrain widely applied is known as Lorenz condition and is written as

$$\nabla.\overrightarrow{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t} = 0, \tag{3.14}$$

where c is the speed of light. $c = \sqrt{\varepsilon\mu}$. ε and μ are the dielectric parameters of the medium, i.e permittivity and permeability of space. Equation (3.3) can be rearranged as

$$\nabla \times \overrightarrow{H} - \frac{\partial \overrightarrow{D}}{\partial t} = \overrightarrow{J}. \tag{3.15}$$

Using (3.11), (3.12), and (3.14) in (3.15) we obtain

$$\nabla \times \left(\nabla \times \overrightarrow{A}\right) - \frac{1}{c^2}\frac{\partial}{\partial t}\left(-\nabla\varphi - \frac{\partial \overrightarrow{A}}{\partial t}\right) = \mu\overrightarrow{J} \tag{3.16}$$

$$\frac{1}{c^2}\frac{\partial^2 \overrightarrow{A}}{\partial t^2} - \nabla^2\overrightarrow{A} = \mu\overrightarrow{J}. \tag{3.17}$$

The expressions in (3.18) and (3.19) constitute the gauge invariance of Maxwell's equations which leaves $\overrightarrow{E}$ and $\overrightarrow{B}$ changed [284] for any scalar function $f\left(\overrightarrow{r}, t\right)$.

$$\varphi' = \varphi - \frac{\partial f}{\partial t} \tag{3.18}$$

$$\overrightarrow{A}' = \overrightarrow{A} + \nabla f. \tag{3.19}$$

Applying (3.18) and (3.19) to (3.14), we can write

$$\nabla.\overrightarrow{A}' + \frac{1}{c^2}\frac{\partial \varphi'}{\partial t} = \nabla.\overrightarrow{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t} - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} + \nabla^2 f. \tag{3.20}$$

Comparing (3.14) and (3.20), we have

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \nabla^2 f = \nabla.\overrightarrow{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t}. \tag{3.21}$$

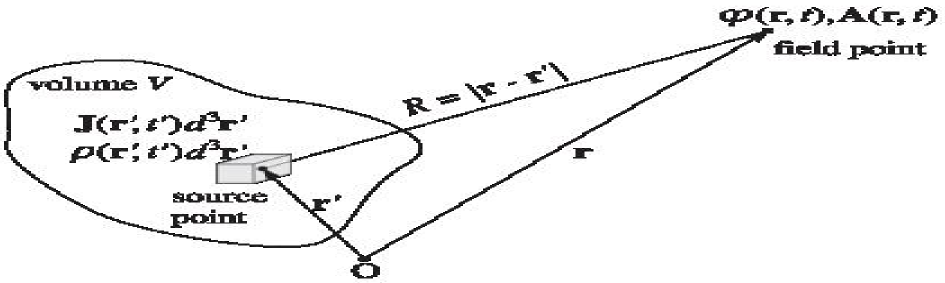


Figure 3.2: Retarded potentials generated by a localised current/charge distribution. Taken from [284].

Equation (3.21) is an inhomogeneous wave equation whose solution f could be used to refine $\overrightarrow{A}$ and φ so that equation (3.14) is satisfied.

Therefore, (3.8) and (3.17) are wave equations for the potentials which are representative of Maxwell's equations. The potentials can be computed if both conduction current density and charge density are known. In each of these two equations, the right hand side represents a continuous source function. In image acquisition using a camera, based on Huygens' principle, discrete functions are considered since any scene object is a set of point sources.

Consider a source point located on a scene object of volume V of known densities as shown in Figure 3.2. On the basis of the principle of causality, it takes $\frac{R}{c}$ seconds for the wave front emanating from the source $\overrightarrow{r}'$ to reach any field points $\overrightarrow{r}$. Hence the potentials computed using (3.8) and (3.17) are referred to as retarded potentials.

Assuming a source point of arbitrary function $f(t)$, (3.8) will take the form

$$\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi = f(t)\,\delta^{(3)}(r)\,, \tag{3.22}$$

where $\delta^{(3)}(r)$ is a three-dimensional 3D delta function. A delta function is defined only for $x = 0$. The value at $x = 0$ is infinity.

Also, assuming the solution of (3.22) at a retarded time $t' = t - \frac{R}{c}$ is

$$\varphi g\left(\overrightarrow{r}, t\right) = \frac{f(t')}{4\pi r} = \frac{f\left(t - \frac{R}{c}\right)}{4\pi r}. \tag{3.23}$$

Using a green's function $g\left(\overrightarrow{r}\right)$ such that

$$g\left(\overrightarrow{r}\right) = \frac{1}{4\pi r}, \tag{3.24}$$

(3.23) becomes

$$\varphi\left(\overrightarrow{r}, t\right) = f\left(t - \frac{r}{c}\right) g\left(\overrightarrow{r}\right) \tag{3.25}$$

and from (3.24) we obtain

$$\nabla g = -\hat{r}\frac{g}{r} \tag{3.26}$$

$$\nabla^2 g = -\delta^{(3)}\left(r\right). \tag{3.27}$$

$\hat{r}$ is a unit vector in the radial direction. The differential of the numerator of (3.23) with respect to r is:

$$\frac{\partial}{\partial r} f\left(t - \frac{r}{c}\right) = -\frac{1}{c}\dot{f}. \tag{3.28}$$

Also,

$$\nabla f = -\hat{r}\frac{\dot{f}}{c} \tag{3.29}$$

$$\nabla^2 f = \frac{1}{c^2}\ddot{f} - \frac{2\dot{f}}{cr}. \tag{3.30}$$

According to [284],

$$\nabla^2 \varphi = \nabla^2\left(fg\right) = 2\nabla f.\nabla g + g\nabla^2 f + f\nabla^2 g \tag{3.31}$$

$$\nabla^2 \varphi = \frac{1}{c^2}\ddot{f}g - f\left(t - \frac{r}{c}\right)\delta^{(3)}\left(\overrightarrow{r}\right). \tag{3.32}$$

From (3.25)

$$\frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} = \frac{1}{c^2}\ddot{f}g, \tag{3.33}$$

(3.32) becomes

$$\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi = f\left(t\right)\delta^{(3)}\left(\overrightarrow{r}\right). \tag{3.34}$$

This means that starting with the causal solution of (3.23) and the use of green's function for electrostatic Coulomb problem, it is possible to derive the corresponding wave equation (3.34) which is the same as (3.22). Therefore, the causal solution of (3.23) is correct.

For another source point at another location with the considered volume of Figure 3.2 $\overrightarrow{r}$ becomes $\overrightarrow{r} - \overrightarrow{r}\prime$. as a result of this change, equation (3.34) becomes

$$\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi = f\left(t\right)\delta^{(3)}\left(\overrightarrow{r} - \overrightarrow{r}\prime\right). \tag{3.35}$$

Therefore the causal solution of (3.35) can be written as

$$\varphi\left(\overrightarrow{r}, t\right) = \frac{f\left(\overrightarrow{r}\prime, \left(t - \frac{R}{c}\right)\right)}{4\pi\left(\overrightarrow{r} - \overrightarrow{r}\prime\right)}. \tag{3.36}$$

Therefore, the total source function due to all the discrete source points in the considered volume can be obtained by performing volume integral. As such we have

$$f\left(\overrightarrow{r}, t\right) = \int\int\int f\left(\overrightarrow{r}, t\right)\delta^{(3)}\left(\overrightarrow{r} - \overrightarrow{r}\prime\right) dxdydz\overrightarrow{r}\prime. \tag{3.37}$$

The potential $\varphi\left(\overrightarrow{r}, t\right)$ corresponding to the source expressed in (3.37) is obtained as

$$\varphi\left(\overrightarrow{r}, t\right) = \int\int\int \frac{f\left(\overrightarrow{r}\prime, \left(t - \frac{R}{c}\right)\right)}{4\pi\left(\overrightarrow{r} - \overrightarrow{r}\prime\right)} dxdydz\overrightarrow{r}\prime. \tag{3.38}$$

In conclusion, (3.38) is the causal solution to the general wave equation expressed in (3.32). Using a similar argument, $\overrightarrow{A}\left(\overrightarrow{r}, t\right)$ can be written as

$$\overrightarrow{A}\left(\overrightarrow{r}, t\right) = \int\int\int \frac{f\left(\overrightarrow{r}\prime, \left(t - \frac{R}{c}\right)\right)}{4\pi\left(\overrightarrow{r} - \overrightarrow{r}\prime\right)} dxdydz\overrightarrow{r}\prime. \tag{3.39}$$

This idea can be applied in order to compute the retarded potentials to the wave equations (3.8) and (3.17) where $f\left(\overrightarrow{r},t\right) = \frac{\rho(\overrightarrow{r},t)}{\varepsilon}$, $f\left(\overrightarrow{r},t\right) = \mu\overrightarrow{J}\left(\overrightarrow{r},t\right)$, and $R = \left|\overrightarrow{r} - \overrightarrow{r}\prime\right|$. That is

$$\varphi\left(\overrightarrow{r},t\right) = \int\int\int \frac{\rho\left(\overrightarrow{r}\prime\right)}{4\pi\varepsilon R} dxdydz\,\overrightarrow{r}\prime, \tag{3.40}$$

$$\overrightarrow{A}\left(\overrightarrow{r},t\right) = \int\int\int \frac{\mu\overrightarrow{J}\left(\overrightarrow{r}\prime\right)}{4\pi R} dxdydz\,\overrightarrow{r}\prime. \tag{3.41}$$

In practice, both the conduction current density $\overrightarrow{J}$ and electric charge density ρ depend on the object to be capture by the camera.

If the source point oscillates with respect to time, then $\varphi\left(\overrightarrow{r},t\right) = \varphi\left(\overrightarrow{r}\right)e^{j\omega t}$, $\overrightarrow{A}\left(\overrightarrow{r},t\right) = \overrightarrow{A}\left(\overrightarrow{r}\right)e^{j\omega t}$, $\rho\left(\overrightarrow{r},t\right) = \rho\left(\overrightarrow{r}\right)e^{j\omega t}$, $\overrightarrow{J}\left(\overrightarrow{r},t\right) = \overrightarrow{J}\left(\overrightarrow{r}\right)e^{j\omega t}$, where ω is the angular frequency of the oscillation. j is the complex component. Therefore, the retarded potentials will take to different forms as

$$\varphi\left(\overrightarrow{r},t\right) = \int\int\int \frac{\rho\left(\overrightarrow{r}\prime\right)e^{j\omega t}e^{j\omega\left(t-\frac{R}{c}\right)}}{4\pi\varepsilon R} dxdydz\,\overrightarrow{r}\prime, \tag{3.42}$$

$$\overrightarrow{A}\left(\overrightarrow{r},t\right) = \int\int\int \frac{\mu\overrightarrow{J}\left(\overrightarrow{r}\prime\right)e^{j\omega t}e^{j\omega\left(t-\frac{R}{c}\right)}}{4\pi R} dxdydz\,\overrightarrow{r}\prime. \tag{3.43}$$

The simplification of (3.42) and (3.43) yield

$$\varphi\left(\overrightarrow{r},t\right) = \int\int\int \frac{\rho\left(\overrightarrow{r}\prime\right)e^{-jkR}}{4\pi\varepsilon R} dxdydz\,\overrightarrow{r}\prime, \tag{3.44}$$

$$\overrightarrow{A}\left(\overrightarrow{r},t\right) = \int\int\int \frac{\mu\overrightarrow{J}\left(\overrightarrow{r}\prime\right)e^{-jkR}}{4\pi R} dxdydz\,\overrightarrow{r}\prime, \tag{3.45}$$

where $k = \frac{\omega}{c}$ is referred to as the wave number.

Therefore, the magnetic and electric fields of (3.5) and (3.6) can be calculated if $\overrightarrow{A}\left(\overrightarrow{r},t\right)$ and $\varphi\left(\overrightarrow{r},t\right)$ are known. However, (3.6) can be made to depend only on vector potential $\overrightarrow{A}\left(\overrightarrow{r},t\right)$ with further simplification. Re-writing Lonrenz condition and recall that $c = \sqrt{\varepsilon\mu}$, and $\frac{\partial}{\partial t} = j\omega$, we can have

$$\nabla.\overrightarrow{A} + j\omega\varepsilon\mu\varphi = 0. \tag{3.46}$$

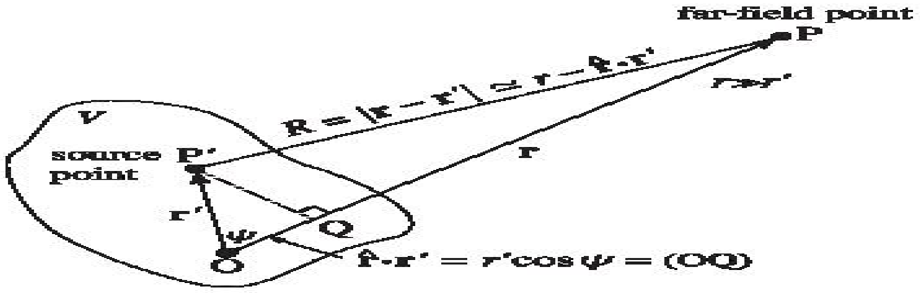


Figure 3.3: Far-field approximation. Taken from [284].

Making φ the subject of the formula, we get

$$\varphi = -\frac{1}{j\omega\varepsilon\mu}\nabla.\overrightarrow{A}. \tag{3.47}$$

Therefore, the electric field expressed in (3.6) becomes

$$\left\{\begin{array}{l} \overrightarrow{E} = \frac{1}{j\omega\varepsilon\mu}\left(\nabla\left(\nabla.\overrightarrow{A}\right) + k^2\overrightarrow{A}\right) \\ \overrightarrow{E} = \frac{1}{j\omega\varepsilon\mu}\left(\nabla\times\left(\nabla\times\overrightarrow{A}\right) - \mu\overrightarrow{J}\right) \end{array}\right\}. \tag{3.48}$$

In electromagnetic imaging problems, interest focuses on the behaviour of the scattered EM wavefield generated by variations in the material parameters ε, μ, and σ. σ is the conductivity of the medium. Also

$$\overrightarrow{H} = \frac{1}{\mu}\nabla\times\overrightarrow{A}. \tag{3.49}$$

In optical imaging, the camera sensor is considered to be located in the fields that have radiated away from their current sources (scene objects). This is because they carry large power. Such far-fields have to satisfy the condition $\overrightarrow{r} \gg \overrightarrow{r}'$ or $\overrightarrow{r} > l$. Where l is the extent of the current distribution in the source. The situation usually obtained in practice and depicted in Figure 3.3 is that at far distances the sides PP' and PQ of the triangle PQP' are almost equal.

Applying cosine rule to triangle PQP' we get

$$R = \left|\overrightarrow{r} - \overrightarrow{r}'\right| = \left(r^2 + r'^2 - 2rr'\cos\psi\right)^{\frac{1}{2}}. \tag{3.50}$$

When r is factored out of the square root, we can write

$$R = r\left(1 - 2\frac{r'}{r}\cos\psi + \frac{r'^2}{r^2}\right)^{\frac{1}{2}} \tag{3.51}$$

Considering the fact that $\overrightarrow{r} \gg \overrightarrow{r}'$ and applying Binomial approximation to (3.51), we have the reduced form as

$$R = r\left(1 - \frac{1}{2}\left(2\frac{r'}{r}\cos\psi\right)\right). \tag{3.52}$$

With further simplification and recognizing that the dot product $\overrightarrow{M}.\overrightarrow{N}$ of two vectors $\overrightarrow{M}$ and $\overrightarrow{N}$ that are positioned with respect to each other at an angle θ between them is $\left|M\right|\left|N\right|\cos\theta$, we get

$$R = r - r'\cos\psi = r - \hat{r}.r' \simeq r. \tag{3.53}$$

We now substitute the approximate form of R (the distance from a volume element dV to the point of observation) in the denominator part of (3.45). The approximation allows for component terms that constitute the waves that are propagated towards the camera sensor. Hence we obtain

$$\overrightarrow{A}\left(\overrightarrow{r},t\right) = \int\int\int \frac{\mu\overrightarrow{J}\left(\overrightarrow{r}'\right)e^{-jk(r-\hat{r}.r')}}{4\pi r} dxdydz\overrightarrow{r}'. \tag{3.54}$$

Rearranging (3.54) we get

$$\overrightarrow{A}\left(\overrightarrow{r},t\right) = \frac{\mu e^{-jkr}}{4\pi r}\int\int\int \overrightarrow{J}\left(\overrightarrow{r}'\right)e^{jk\hat{r}.r'}dxdydz\overrightarrow{r}'. \tag{3.55}$$

To use the camera aperture as field source of the radiated fields, a variation of the Huygens-Fresnel principle needs to be used. This principle states that the points on each wave-front become the sources of secondary spherical waves propagating outwards and whose superposition generates the next wave-front. Therefore, two types of aperture will be considered: Rectangular and circular.

3.3.1 *Rectangular Aperture*

In [285–287], it is thought that an aperture can be a source. We are now interested in the field distribution emanating from it. If we consider an infinitesimally small

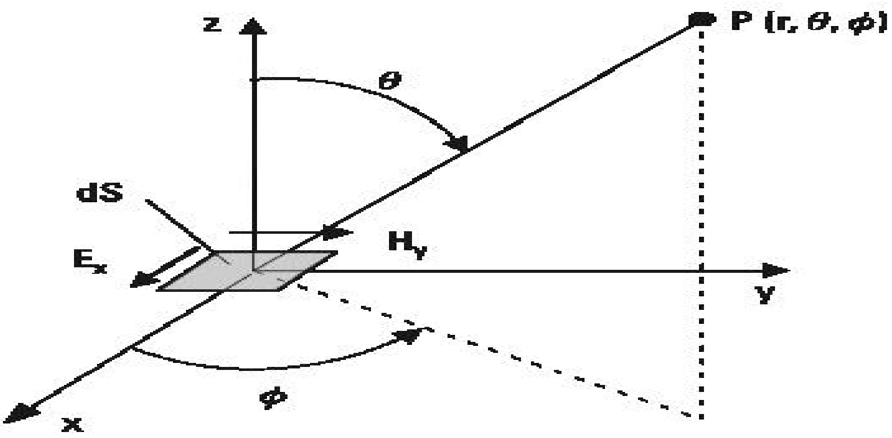


Figure 3.4: Radiating surface element

camera aperture (pinhole) as shown in Figure 3.4 to be a source of surface dS, with the surface current density distribution J_s, then (3.55) can be written as

$$d\overrightarrow{A}\left(\overrightarrow{r},t\right) = \frac{\mu e^{-jkr}}{4\pi r} \int\int \overrightarrow{J_s}\left(\overrightarrow{r}'\right) e^{jk\hat{r}.r'} dS\overrightarrow{r}'. \tag{3.56}$$

$\overrightarrow{J_s}$ is expressed as

$$\hat{J_s} = \hat{n} \times \overrightarrow{H_y} = -\frac{E_x}{\eta} r', \tag{3.57}$$

where η is the wave impedance. E_x and H_y are the electric and magnetic fields in the x and y directions.

The wave exiting the aperture is spherical. In the region farther from the aperture, the field distribution can be approximated as parabolic wave. This is referred to as the Fresnel region. As the waves travel even further away from the aperture the spherical waves become approximately plane waves. Usually the camera lens will bring the propagating light in the far-field (Fraunhofer diffraction region) to a focus on the sensor to form the image.

To model the far-field, the solutions of Maxwell's equations and the directional patterns of a source are best described in spherical coordinates since the fields radiated by sources of finite dimensions are spherical. Therefore, the magnetic vector potential is expressed in component form as

$$\left\{\begin{array}{c} A_r = \sin\theta\cos\phi A_x \\ A_\theta = \cos\theta\cos\phi A_x \\ A_\phi = -\sin\phi A_x \end{array}\right\} \tag{3.58}$$

$$\begin{aligned}\nabla\times\overrightarrow{A} = \hat{r}\frac{1}{r\sin\theta}\left(\frac{\partial\left(\sin\theta A_\phi\right)}{\partial\theta} - \frac{\partial A_\theta}{\partial\phi}\right) \\ +\hat{\theta}\frac{1}{r}\left(\frac{1}{\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{\partial\left(rA_\phi\right)}{\partial r}\right) \\ +\hat{\phi}\frac{1}{r}\left(\frac{\partial\left(rA_\theta\right)}{\partial r} - \frac{\partial A_r}{\partial r}\right)\end{aligned} \tag{3.59}$$

$$\nabla\times\overrightarrow{A} = -\frac{\partial A_\phi}{\partial r}\hat{\theta} + \frac{\partial A_\theta}{\partial r}. \tag{3.60}$$

Using (3.48) under the assumption of free-space (points removed from source, $\hat{J} = 0$), [288] the electric field components in the θ and ϕ directions can be obtained as

$$E_\theta = \frac{jE_x e^{-jkr}}{2\lambda r}\left(\cos\theta + 1\right)\cos\phi dS \tag{3.61}$$

$$E_\theta = -\sin\phi\frac{jE_x e^{-jkr}}{2\lambda r}\left(\cos\theta + 1\right)dS \tag{3.62}$$

$$H_\theta = -\frac{E_\phi}{\eta} \qquad H_\phi = \frac{E_\theta}{\eta} \qquad H_r = 0 \qquad E_r = 0. \tag{3.63}$$

For a camera aperture of dimensions a and b in the x-y plane shown in Figure 3.5, the radiated electric is the summation of all the contributions by the infinitesimally small sources within the dimension of the aperture. Therefore, we can write

$$E_\theta = \frac{jE_x e^{-jkr}}{2\lambda r}\left(\cos\theta + 1\right)\int\limits_{-m}^{m}\int\limits_{-n}^{n} e^{-jk\hat{r}.r'}dx'dy', \tag{3.64}$$

where $m = \frac{b}{2}$, $n = \frac{a}{2}$, and

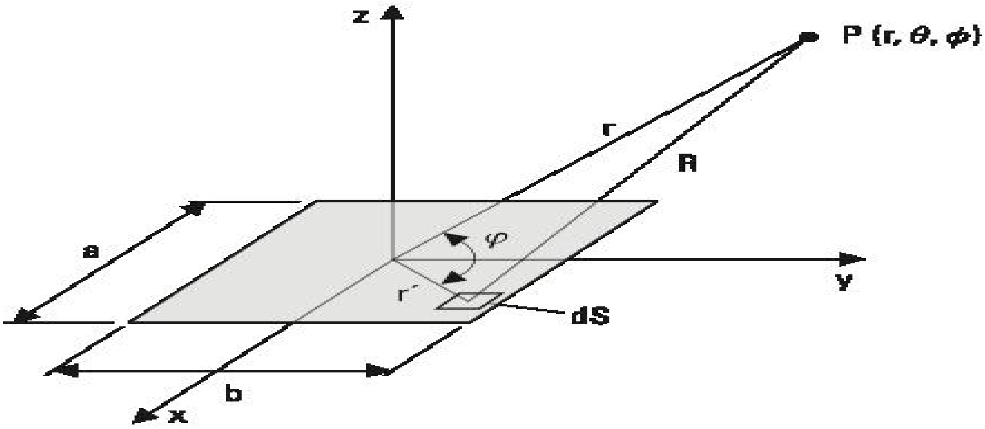


Figure 3.5: Radiation aperture of a by b dimension.

$$\hat{r}.r' = (\cos\phi \sin\theta)\, x' + (\sin\phi \sin\theta)\, y'. \tag{3.65}$$

Therefore, (3.64) becomes

$$E_\theta = \left(\frac{\sin\left(\frac{1}{2}kb\sin\theta\sin\phi\right)}{\frac{1}{2}kb\sin\theta\sin\phi} \right) \left(\frac{\sin\left(\frac{1}{2}ka\sin\theta\cos\phi\right)}{\frac{1}{2}ka\sin\theta\cos\phi} \right) \sin\phi \\ (1+\cos\theta)\,\frac{jabE_x}{2\lambda r} e^{-jkr}, \tag{3.66}$$

where λ is the wavelength. Similarly

$$E_\phi = \left(\frac{\sin\left(\frac{1}{2}kb\sin\theta\sin\phi\right)}{\frac{1}{2}kb\sin\theta\sin\phi} \right) \left(\frac{\sin\left(\frac{1}{2}ka\sin\theta\cos\phi\right)}{\frac{1}{2}ka\sin\theta\cos\phi} \right) \cos\phi \\ (1+\cos\theta)\,\frac{jabE_x}{2\lambda r} e^{-jkr}. \tag{3.67}$$

According to [285], E_θ and E_ϕ only exist in the yz-plane $\left(\phi = \frac{\pi}{2}\right)$ and xz-plane $(\phi = 0)$ respectively. It is further stated that significant field is generated at small angles only if the aperture is large such that a and $b \gg \lambda$. This means $\cos\theta \approx 1$ and $\sin\theta = \theta$. Therefore, the field pattern yz-plane $(E - plane)$ becomes

$$E_r = E_\phi = 0,\ E_\theta = \left(\frac{\sin\left(\frac{1}{2}kb\theta\right)}{\frac{1}{2}kb\theta} \right) \frac{jabE_x}{\lambda r} e^{-jkr} \tag{3.68}$$

For the xz-plane, that is (H – plane), we have

$$E_r = E_\theta = 0,\ E_\phi = \left(\frac{\sin\left(\frac{1}{2}ka\theta\right)}{\frac{1}{2}ka\theta}\right)\frac{jabE_x}{\lambda r}e^{-jkr}. \tag{3.69}$$

Equation (3.68) and (3.69) describe the three-dimensional 3D electric field distribution that is focused on the camera sensor. A 3D plot of the field patterns can be seen in Figure 3.6. It can be observed that because the dimensions of the aperture is one wavelength, there are no sidelobes. However, when the aperture dimensions are a multiple of one wavelength, multiple sidelobes begin to appear. The reason for this observation can be found in (3.68) and (3.69). The E-plane field distribution is a function of the dimension b of the aperture while for the H-plane it is dependent on a. Therefore, the number of sidelobes increases as the aperture dimension increases [288].

Also plots of the field strength against theta θ, for some aperture dimensions are given in Figure 3.7 and Figure 3.8.

Therefore, the intensity distribution $I\left(x,y\right)$ recorded on the camera sensor is

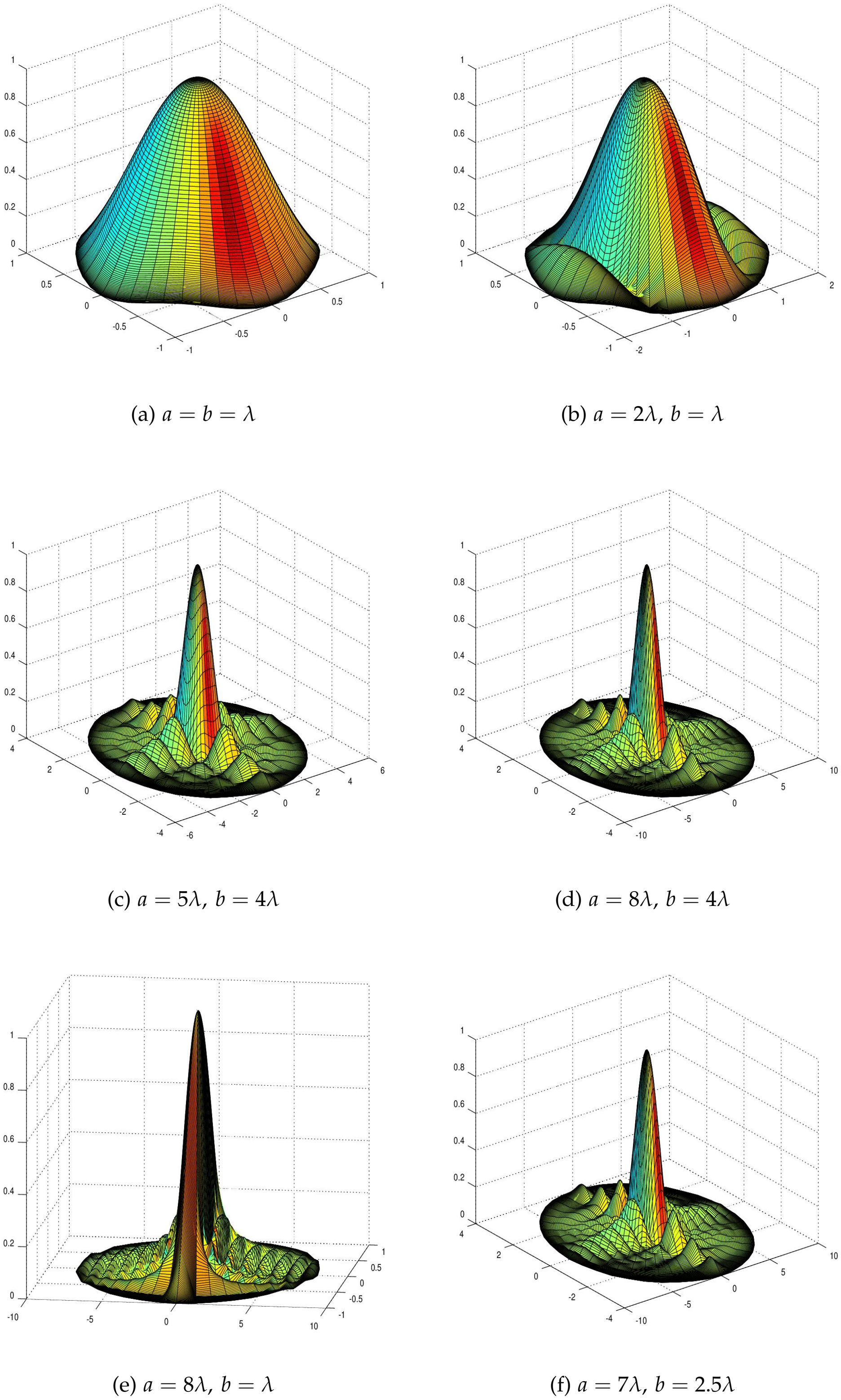
$$I\left(x,y\right) = \left|E_\theta\right|^2 = \frac{E_x^2}{\left(\lambda r\right)^2}\left|ab\left(\left(\frac{\sin\left(\frac{1}{2}ka\theta\right)}{\frac{1}{2}ka\theta}\right)^2 + \left(\frac{\sin\left(\frac{1}{2}kb\theta\right)}{\frac{1}{2}kb\theta}\right)^2\right)\right|^2. \tag{3.70}$$

Equation (3.70) expresses the diffraction ability of the aperture. the consequence of this is that different scene points are made to spread out.

Experiment has shown that because camera sensors are not perfect, a scene point will always be blurred as shown in Figure 3.9. If a lens produce the same blurring effect irrespective of the position of the scene point, then the lens is said to be linear shift-invariant. Therefore, the blur point can be considered as the point Point Spread Function (PSF) of the lens. The challenge now is how to obtain this **PST!** (**PST!**).

From (3.70), it is observed that the $I\left(x,y\right)$ is directly proportional to PSF. We extract PSF as

$$PSF = ab\left|\left(\left(\frac{\sin\left(\frac{1}{2}ka\theta\right)}{\frac{1}{2}ka\theta}\right)^2 + \left(\frac{\sin\left(\frac{1}{2}kb\theta\right)}{\frac{1}{2}kb\theta}\right)^2\right)\right|^2. \tag{3.71}$$



(a) $a = b = \lambda$

(b) $a = 2\lambda$, $b = \lambda$

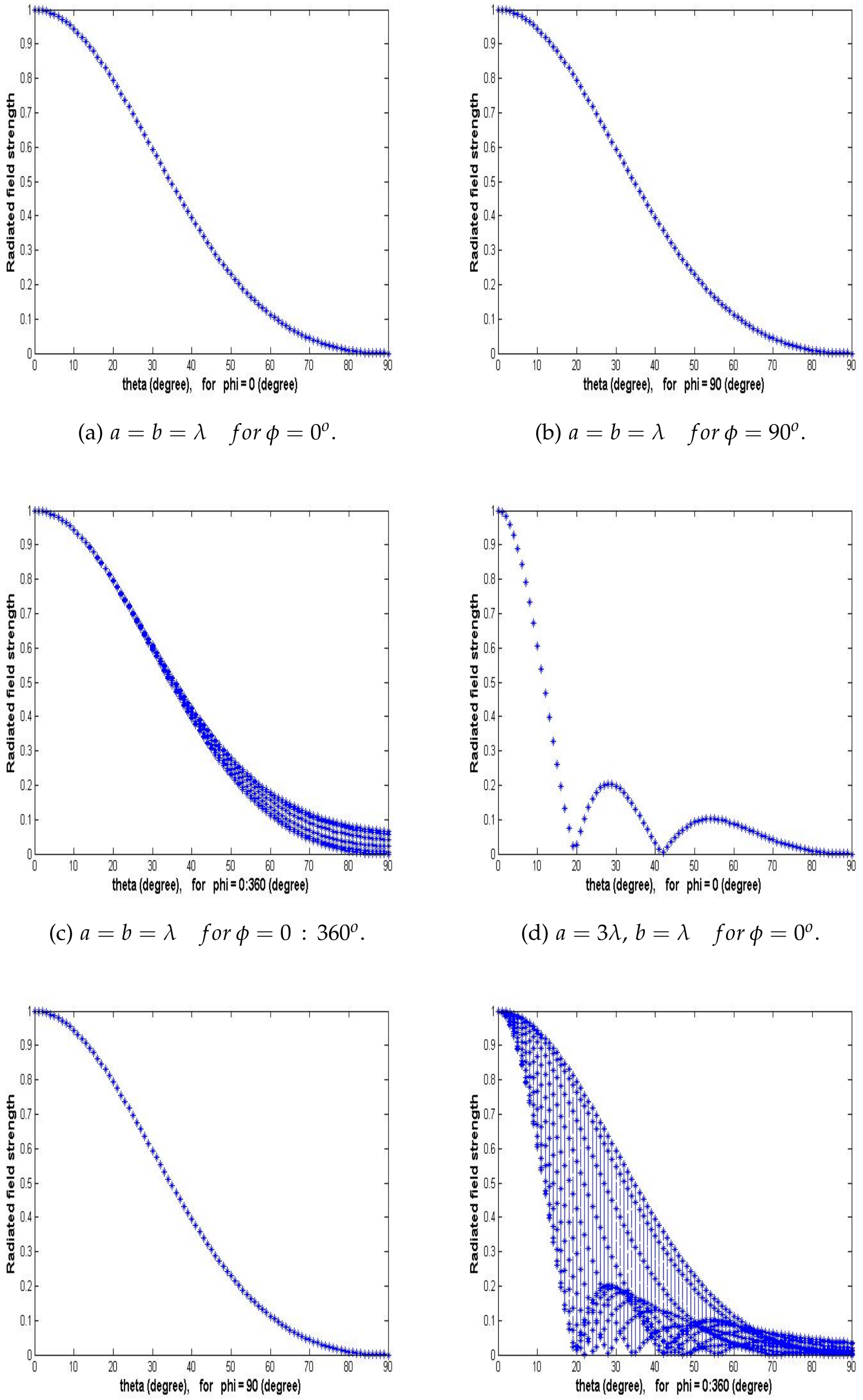
(c) $a = 5\lambda$, $b = 4\lambda$

(d) $a = 8\lambda$, $b = 4\lambda$

(e) $a = 8\lambda$, $b = \lambda$

(f) $a = 7\lambda$, $b = 2.5\lambda$

Figure 3.6: 3D plot of radiation pattern focused on camera sensor based on the use of rectangular aperture.



(a) $a = b = \lambda \quad for\, \phi = 0^o$.

(b) $a = b = \lambda \quad for\, \phi = 90^o$.

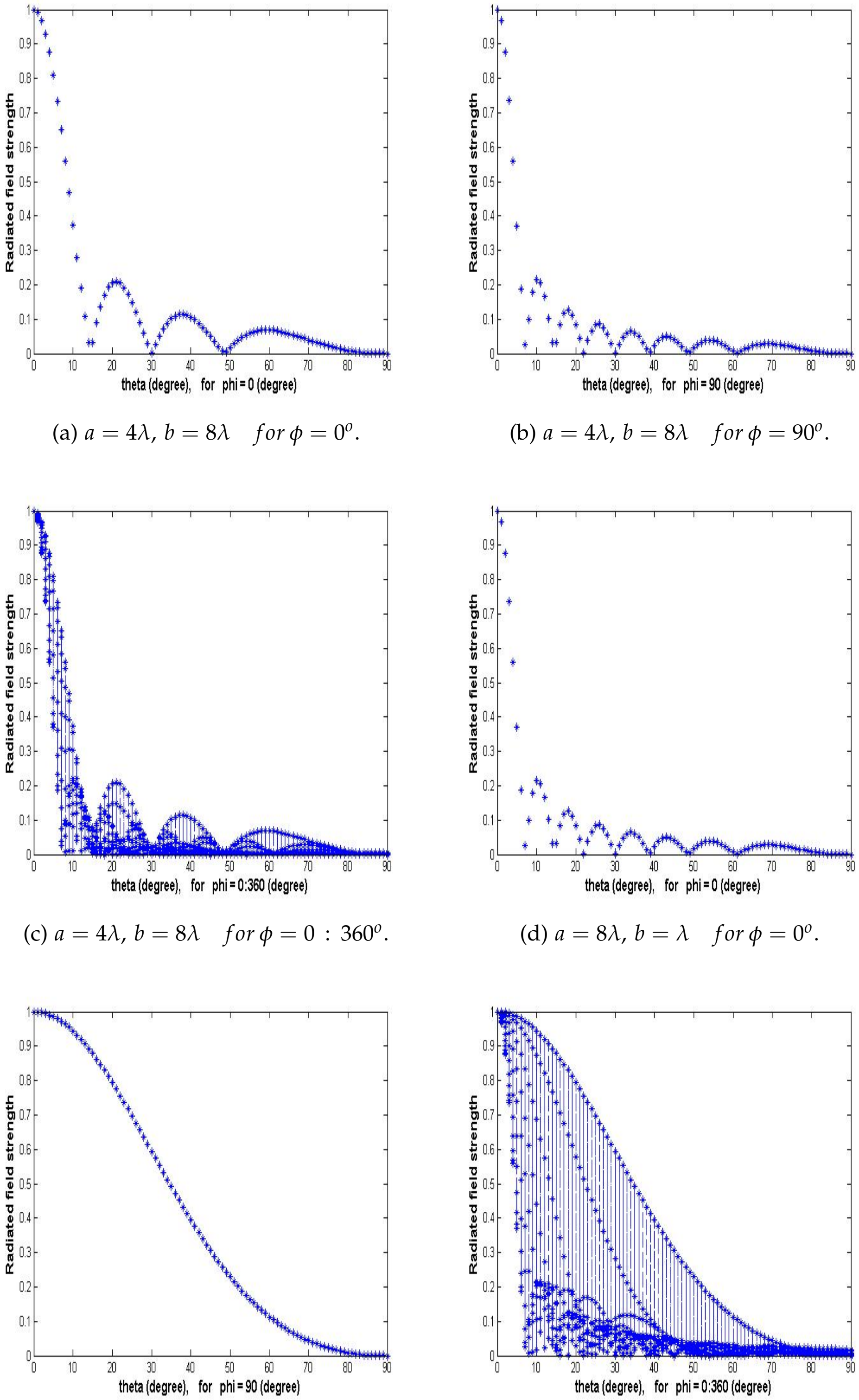
(c) $a = b = \lambda \quad for\, \phi = 0 : 360^o$.

(d) $a = 3\lambda,\, b = \lambda \quad for\, \phi = 0^o$.

(e) $a = 3\lambda,\, b = \lambda \quad for\, \phi = 90^o$.

(f) $a = 3\lambda,\, b = \lambda \quad for\, \phi = 0 : 360^o$.

Figure 3.7: 2D plot of field strength against theta.



(a) $a = 4\lambda$, $b = 8\lambda \quad for\, \phi = 0^o$.

(b) $a = 4\lambda$, $b = 8\lambda \quad for\, \phi = 90^o$.

(c) $a = 4\lambda$, $b = 8\lambda \quad for\, \phi = 0\ :\ 360^o$.

(d) $a = 8\lambda$, $b = \lambda \quad for\, \phi = 0^o$.

(e) $a = 8\lambda$, $b = \lambda \quad for\, \phi = 90^o$.

(f) $a = 8\lambda$, $b = \lambda \quad for\, \phi = 0\ :\ 360^o$.

Figure 3.8: 2D plot of field strength against theta.

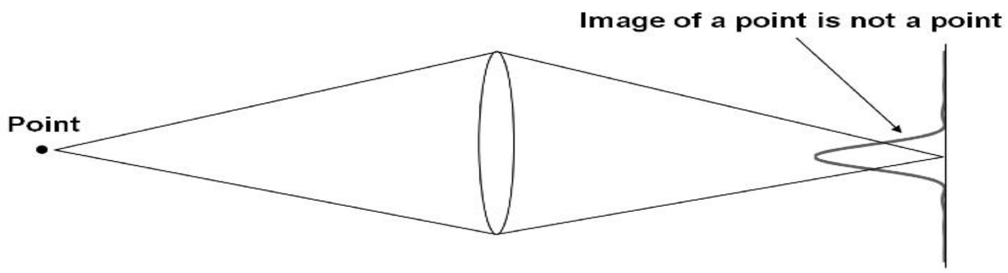


Figure 3.9: The image of a scene point is not a perfect point on the camera sensor. Taken from [279].

Consider a scene as a set of points with different intensity value represented by $p\left(x,y\right)$. If the image of $p\left(x,y\right)$ is $P\left(x,y\right)$, then the image formation process can be expressed in convolution form as

$$P\left(x,y\right) = PSF\left(x,y\right) * p\left(x,y\right). \tag{3.72}$$

Writing (3.72) in frequency domain we have

$$Fouriertransform[P\left(x,y\right)] = Fouriertransform[PSF\left(x,y\right) * p\left(x,y\right)]. \tag{3.73}$$

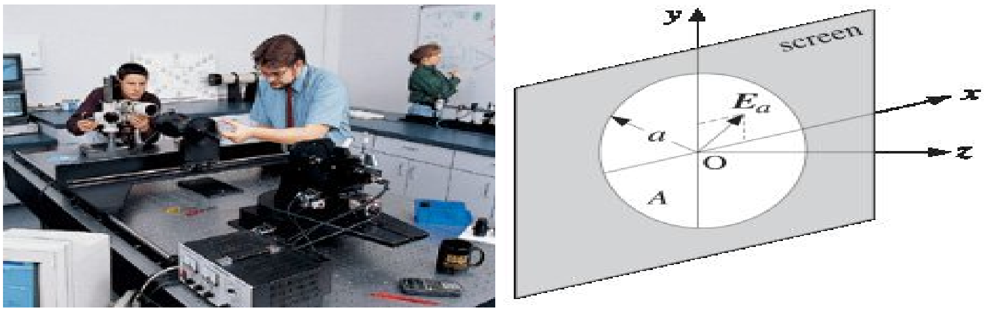
Hence the Fourier transform of PSF is the Optical Transfer Function (OTF) of the camera. It is expressed as

$$OTF = Fouriertransform[PSF\left(x,y\right)]. \tag{3.74}$$

In the normalised form, the OTF is referred to as MTF. This is an important parameter used for performance characterisation of cameras and image quality as shown in Figure 3.10 (a). Further discussion on MTF can be found in the appendix section. This parameter will be used in the objective assessment of some of the output images in this thesis.

3.3.2 *Circular Aperture*

The analysis, characterisation, and design of circular camera aperture has received considerable research attention. This configuration can easily be construc-



(a)

(b)

Figure 3.10: (a) Testing of cameras: http://www.edmundoptics.com/technical-resources-center/, (b) circular aperture

ted. The formulation of the problem of determining the field distribution exiting a circular aperture will involve the adoption of a cylindrical coordinate system. This means

$$\hat{r}.r' = \rho' \sin\theta \cos\phi' dS' = \rho' d\rho' d\phi' \tag{3.75}$$

For the circular aperture of Figure 3.10 (b), the field distribution will take the form

$$E_\theta = \sin\phi \frac{jka^2 E_x e^{-jkr}}{2\pi r} \int\limits_0^a \int\limits_0^{2\pi} e^{jk\rho' \sin\theta \cos\phi'} \rho' d\rho' d\phi' \tag{3.76}$$

simplifying and rearranging (3.76), we get

$$E_\theta = 2j \sin\phi \frac{ka^2 E_x e^{-jkr}}{r} \int\limits_0^a \left(\frac{1}{2\pi} \int\limits_0^{2\pi} e^{jm\cos\phi'} d\phi' \right) \rho' d\rho' \tag{3.77}$$

where $m = k\rho' \sin\theta$. Also $\frac{1}{2\pi} \int_0^{2\pi} e^{jm\cos\phi' d\phi'}$ can be represented with Bessel function $J_0(m)$. i.e

$$J_0(m) = \frac{1}{2\pi} \int\limits_0^{2\pi} e^{jm\cos\phi' d\phi'} \tag{3.78}$$

Equation (3.77) becomes

$$E_\theta = 2j \sin\phi \frac{ka^2 E_x e^{-jkr}}{r} \int\limits_0^a J_0(k\rho' \sin\theta) \rho' d\rho' \tag{3.79}$$

Therefore, (3.79) can be written in the absolute and normalised form as

$$\left|E_\theta\right| = 2\frac{J_1\left(ak\sin\theta\right)}{ak\sin\theta} \tag{3.80}$$

Writing (3.80) in terms of wavelength, we have

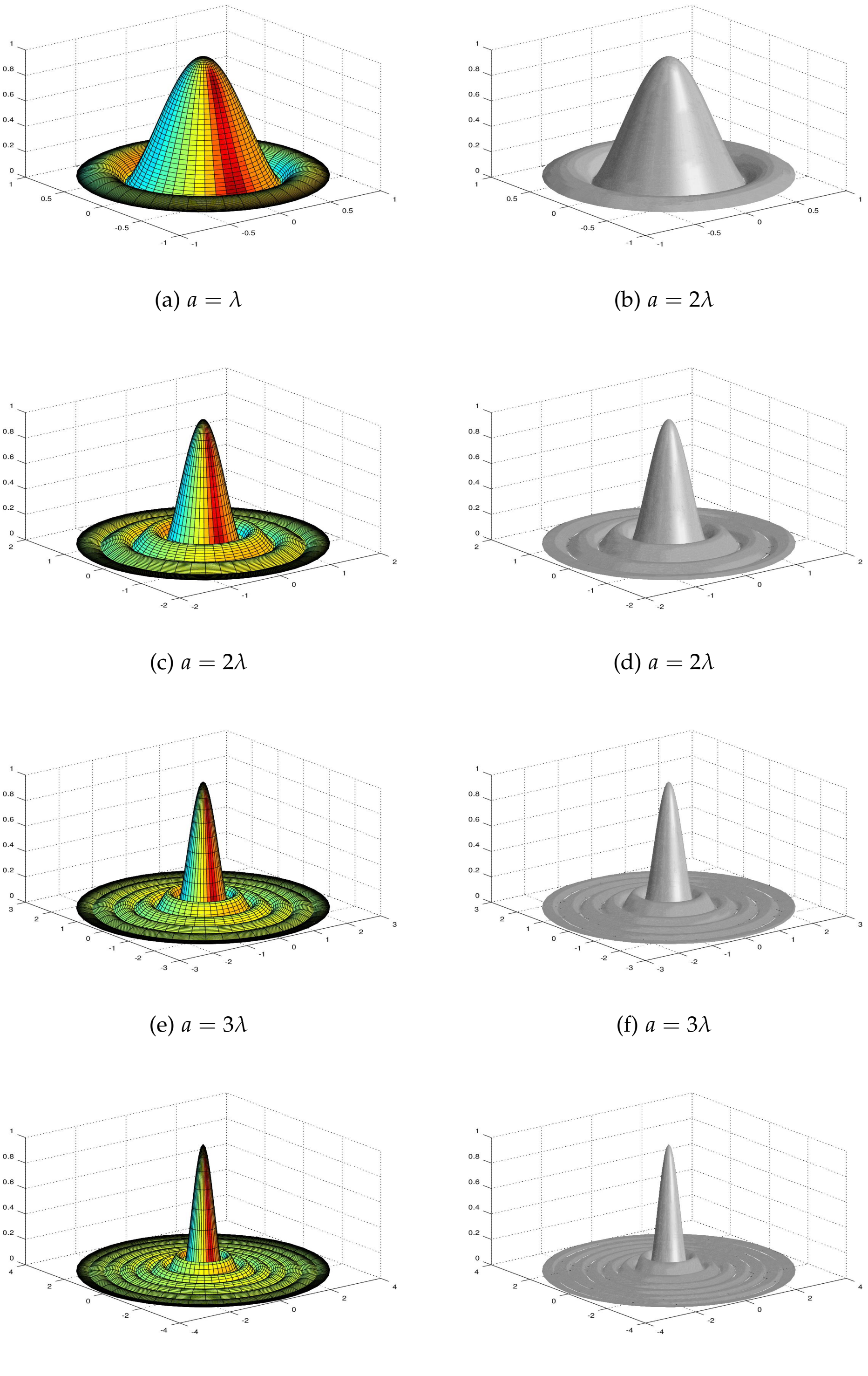
$$\left|E_\theta\right| = 2\frac{J_1\left(2\pi * \frac{1}{2\pi} * a * \frac{2\pi}{\lambda} * \sin\theta\right)}{2\pi * \frac{1}{2\pi} * a * \frac{2\pi}{\lambda} * \sin\theta} = 2\frac{J_1\left(2\pi\frac{a}{\pi\sin\theta}\right)}{2\pi\frac{a}{\pi\sin\theta}} \tag{3.81}$$

Equation (3.81) is true since $k = \frac{2\pi}{\lambda}$. 3D field patterns for circular aperture are shown in Figure 3.11.

A 2D plot of radiated field strength against the angle θ is shown in Figure 3.12.

3.4 CHAPTER SUMMARY

Modelling goes beyond the pinhole perspective model. There exist many other types of simple camera models that are often used for modelling various imaging systems under different practical conditions. An important camera and image quality parameter known as MTF is the basis of the modelling presented in this chapter. It is experimentally observed that the number of sidelobes increases as the aperture dimension increases. The sidelobe is a measure of the spread out of an imaged point. Therefore, a high quality image of a point would be produced when the camera aperture is one wavelength. A measure of the ability, in frequency domain, of a camera to produce an exact replica of a point is referred to as MTF. Consequently, this reflects in the overall quality of the image of an object which is usually a set of points. Also, MTF is use for objective quality assessment of both cameras and images.



(a) $a = \lambda$

(b) $a = 2\lambda$

(c) $a = 2\lambda$

(d) $a = 2\lambda$

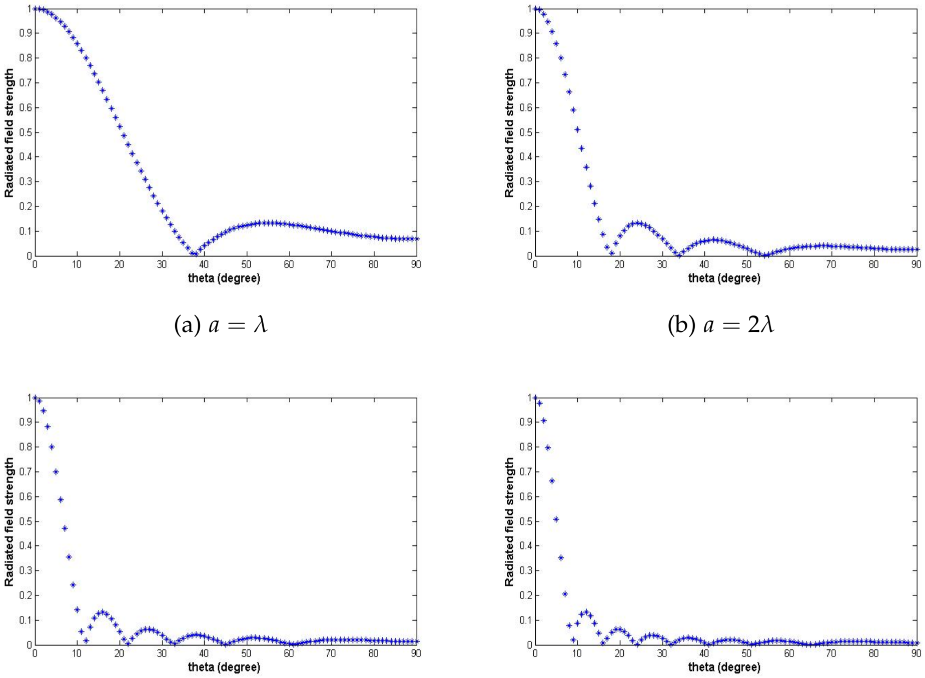
(e) $a = 3\lambda$

(f) $a = 3\lambda$

(g) $a = 4\lambda$

(h) $a = 2\lambda$

Figure 3.11: 3D radiated field pattern of circular aperture.



(a) $a = \lambda$

(b) $a = 2\lambda$

(c) $a = 3\lambda$

(d) $a = 4\lambda$

Figure 3.12: 2D radiated field pattern of circular aperture.

Part III

RESEARCH CONTRIBUTION

4 VIRTUAL CAMERA REALISATION

4.1 GENERAL INTRODUCTION

The minimisation of the number of physical cameras used in a multi-view video set-up has been advocated and actively researched. One important way this can be achieved is through the use of virtual cameras. A virtual camera can be realised by interpolating the intensity value of the pixels contained in the images acquired by view number of physical cameras. Therefore, this chapter discusses the implementation of virtual camera. Its realisation is possible with IBR techniques. These are IBR with and without geometry. To do this, the work in this chapter is presented in three sections. In section 4.2, rendering of image frame is based on the generation of good quality depth map. The generation of 3D panorama is pursued in section 4.3 by using large number of image samples in the creation of stereo panoramic image and their subsequent superposition. Section 4.4 presents the creation of a virtual environment from a pair of panoramic images through depth map texturing. The results were obtained based on the implementations in Matlab environment. In each of the three sections the processes of feature detection, extraction, matching and the computation of homography have been performed. Correction of radial distortion, motion model calculation, mosaic compositing , and blending are additional processes which have been performed in section 4.3.

4.2 METRIC ASPECT OF IMAGE-BASED RENDERING

4.2.1 *Introduction*

Depth image-based rendering has gained significant acceptance having been at the frontiers of understanding and realisation of 3D and free-view televisions. The existence of holes on a rendered image and the occurrence of depth discontinuity on the surface of the object at virtual image which forms the basis of view-dependent depth can be significantly minimised if stringent considerations are exercised on region match measures such as sum of absolute differences, Sum of square differences, and normalised cross-correlation used during interest point detection and feature matching. This study seeks an understanding of the role played by region match measures. These measures affect the interest points used in the generation of depth map which consequently affects the quality of textured depth map of a virtual viewpoint in between two reference images obtained from existing cameras. The understanding gained in this section is used in the selection of region match measure during feature detection and extraction process in the works presented in this thesis.

4.2.1.1 *Pin-Hole Camera Model and Coordinates of Virtual Viewpoint*

We consider a circumstance where the distance between cameras relative to the viewed scene is significant (a "wide" baseline). Also, the epipolar geometry of the scene is unknown and has to be determined using point correspondences. One camera may also go through a significant rotation and translation. From Figure 4.1, [190], it is obvious that if we can locate the same world point in another image taken from different but known pose, we can determine another ray along which that world point must lie. The world point lies at the intersection of these two rays – a process known as triangulation or 3D reconstruction. Even more significantly, if we observe sufficient points, we can estimate the 3D motion of the camera between the views as well as the 3D structure of the world. We define linear transformation in homogeneous coordinates with respect to camera reference frame as the projective transformation from $G \epsilon S^n$ to $g \epsilon R^m$ as

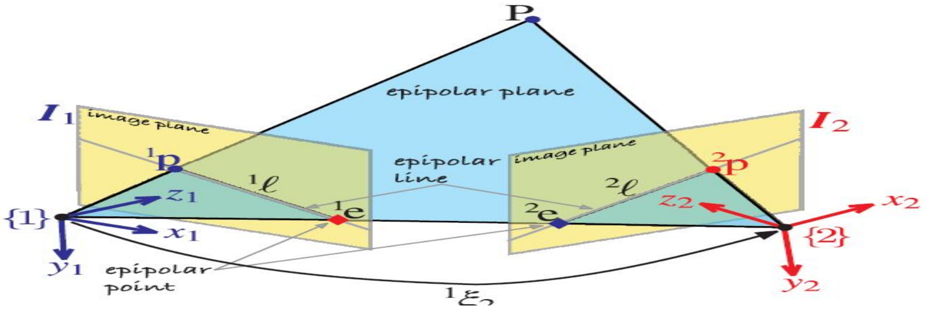


Figure 4.1: Two view camera geometry.

$$x = TP. \tag{4.1}$$

T is a $(m+1) \times (n+1)$ matrix of full rank while x and P are represented in homogeneous coordinates for S and R. For a transformation of 3D world point to a 2D image point $m = 2$ and $n = 3$. T is called camera projection matrix. Let T_1 and T_2, be the projection matrices of camera at point {1} and {2} in Figure 4.1. For the two camera positions {1} and {2} the following projective transformation is true

$$x_1 = T_1 P = K_1 \left[R_1 | t_1\right] P = \left[B_1 | b_1\right] P, \tag{4.2}$$

$$x_2 = T_2 P = K_2 \left[R_2 | t_2\right] P = \left[B_2 | b_2\right] P. \tag{4.3}$$

Camera centres, [71], $C1$ and $C2$ are obtained as

$$C1 = -B_1^{-1} * b_1. \tag{4.4}$$

$$C2 = -B_2^{-1} * b_2. \tag{4.5}$$

Where p is a 3D world point, K_1 and K_2 are the camera calibration matrices which encode the transformation in the image plane from the so-called normalised camera coordinates to pixel coordinates, R_1 and R_2 are the rotations, t_1

and t_2 are translations of camera {1} and {2} respectively. Based on the camera geometry of Figure 4.1, the vectored camera positions {1} and {2} are $C1$ and $C2$ respectively [190]. The distance $C1C2$, between the two cameras is expressed in (4.6). Interpolation can be performed using (4.7) to determine the vectored coordinate $\left[C_{iv}, C_{jv}, C_{kv}\right]$, of any $(n-1)$ virtual view points, c_v along the line linking the two reference cameras

$$\begin{aligned}(C1C2)^2 = 2 & \\ & -2 * \left(\left(C1\left(1,1\right)/c1\right) * \left(C2\left(1,1\right)/c2\right)\right) \\ & -2 * \left(\left(C1\left(2,1\right)/c1\right) * \left(C2\left(2,1\right)/c2\right)\right) \\ & \qquad -2 * \left(\left(C1\left(3,1\right)/c1\right) * \left(C2\left(3,1\right)/c2\right)\right), \end{aligned} \tag{4.6}$$

where

$$c_1 = \sqrt{\left(C1\left(1,1\right)^2 + C1\left(2,1\right)^2 + C1\left(3,1\right)^2\right)}$$

$$c_2 = \sqrt{\left(C2\left(1,1\right)^2 + C2\left(2,1\right)^2 + C3\left(3,1\right)^2\right)}$$

$$C_v = \begin{bmatrix} C_{iv} \\ C_{jv} \\ C_{kv} \end{bmatrix} = \begin{bmatrix} C1\left(1,1\right) : \frac{C2(1,1)-C1(1,1)}{n} : C2\left(1,1\right) \\ C1\left(2,1\right) : \frac{C2(2,1)-C1(2,1)}{n} : C2\left(2,1\right) \\ C1\left(3,1\right) : \frac{C2(3,1)-C1(3,1)}{n} : C2\left(3,1\right) \end{bmatrix}, \tag{4.7}$$

where $C_1 \leq C_v \leq C2$.

The virtual viewpoint image, m_v, which is an array of x_v, is obtained using (4.8). The parameter β represents the distance of P from the camera projection centre and is obtained from the depth map.

$$m_v = \beta K_2 R_2 K_1^{-1} m_1 + K_2 C_v. \tag{4.8}$$

4.2.2 *Point Correspondence and Image Rendering*

It has been realized that the assumption of known camera focal length and principal point in the non-iterative algorithm to solve the problem of relative camera placement [289] may not necessarily be true in certain circumstances. Hence, attention is being shifted to the use of cameras in which the so-called intrinsic parameters are not known. These are referred to as uncalibrated cameras. The use of this class of cameras for multi-view geometry and image rendering demands that feature correspondence, extraction, and matching from a given number of reference images be done accurately. In this work, we start with two reference images obtained from uncalibrated cameras of the same type and model.

4.2.2.1 *Interest Point Detector and Descriptor*

Apart from the non-scale-invariant Harris interest point detector which is based on the eigenvalues of the second-moment matrix, Harris-Laplace and Hessian-Laplace is another corner detector with a predictably receptive acceptance [290]. Issues of changes in longer viewpoint, maximisation of entropy within a region, and speed have all been considered under propositions such as affine-invariant feature detectors, edge-based region detector, and Difference of Gaussians (DoG) filter [291]. The practical adaptability and speed of Scale-Invariant Feature Transform (SIFT) has given it an edge over other point descriptors. This approach has a special fascination in the sense that scale-space features are detected and characterized in a manner invariant to location, scale and orientation. The local image region is represented with multiple images representing each of a number of orientation planes. The effect of this is that variation in local geometry is submerged. In the opinion of [292], this results in an unnecessarily high dimensional SIFT key vector which is not truly invariant to affine distortions.

In this work, a mix of Speeded Up Robust Features (SURF) and corner feature algorithms characterised by repeatability, distinctiveness, and robustness is used for feature detection. SURF is a combination of fast Fast-Hessian detector and a descriptor based on distribution of Haar-wavelet responses within the interest point neighbourhood. Both the detector and descriptor make use of intensity images to achieve speed performance. Metric threshold of 200, Number octaves of 1, and number scale levels of 4 have been used in feature detection. In the feature extraction process, the descriptor computes a histogram of local oriented gradients around the interest point and stores the bins in a 128-dimensional vector.

4.2.2.2 *Feature Matching*

Comparison of two descriptor vectors of invariants leads to efficient computation of similarity factors between two interest points from different images. The similarity factor is used to determine a set of corresponding points from the interest points in two images. Three region match measures are considered in this work namely SAD, SSD, and NCC.

Geometric constraint is used to remove any outliers from the set of corresponding points to find a reliable set of matches that can lead to successful implementation of statistical method, Random Sampling Consensus (RANSAC) algorithm [293]. The outcome of this process is the generation of both fundamental matrix and inliers. Their reliability is confirmed by application of epipolar constraint which requires that the epipoles lie outside (infinity) any of the two reference images.

4.2.2.3 *Dense Stereo Disparity Image*

Feature points are extracted from a pair of two images. This is followed by the establishment of correspondence points. The imposition of epipolar constraint is performed to remove outliers. Rectification transformations are then determined based on the obtained inliers. The two reference images are rotated to obtain a rectified image pair with parallel optical axes where point correspondences between the two camera images always lie on the same epipolar line. These rotations are carried out using rectification homographs generated from the fundamental matrix which is an output of RANSAC algorithm. As a consequence, the search of correspondences is limited to the horizontal direction. Clearly, this eases subsequent disparity matching process.

The generation of a disparity map is generally based on matching cost computation, cost (support) aggregation, disparity computation and optimization, and disparity refinement [294]. In this work, the generated disparity map is followed up with morphological operation to fill any present in the disparity map.

4.2.2.4 *Image Rendering*

- A virtual view point coordinate is computed using (4.7).
- The range depth map and the left reference image are used in (4.8) to render a novel view at a specified virtual view point coordinate.
- Only virtual viewpoints which lie on the line defined by $C1$ and $C2$ are considered in this work ("interpolation"). Although, extrapolation is also possible to some degree.

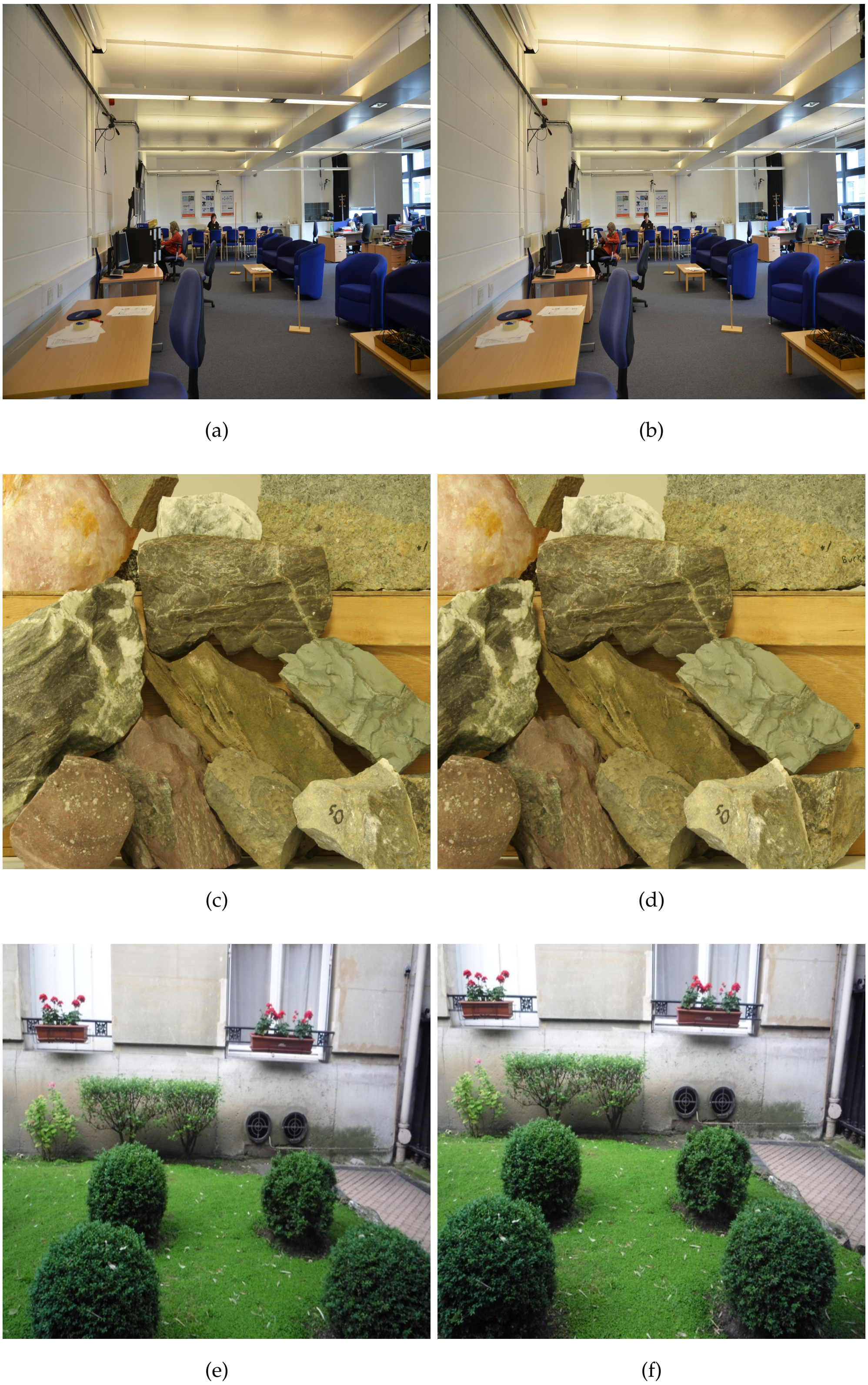
4.2.3 *Experimental Results and Discussion*

4.2.3.1 *Correspondence Points*

Figure 4.2 shows three sets of reference images each containing a left and right component that have been used in our work. In Figure 4.3, it is seen that no interest point is detected when homogeneous image regions such as dark shadows, smooth surface, or wall is encountered during the search process. However, NCC, SSD, and SAD region match measures have performed better on sets (a), (b), and (c) of Figure 4.2 respectively. NCC has an edge over SAD and SSD when depth of the scene is significant such as in Figure 4.2(a). It is observed that in Figure 4.4 all the outliers have been removed using geometric and epipolar constraints and that the corresponding points are in the same rows demonstrating that the rectification was successful.

4.2.3.2 *Depth map*

Depth maps for the three sets of reference images are shown in Figure 4.6. The abnormal spots which are more in (a) and (c) compared with (b) can be attributed to the scene depth. The depth in (b) is shallow. These spots can be removed by using closed and open morphological operators with an appropriate structuring element. In Figure 4.6 (a), the depth increases from the front of the figure towards



(a)

(b)

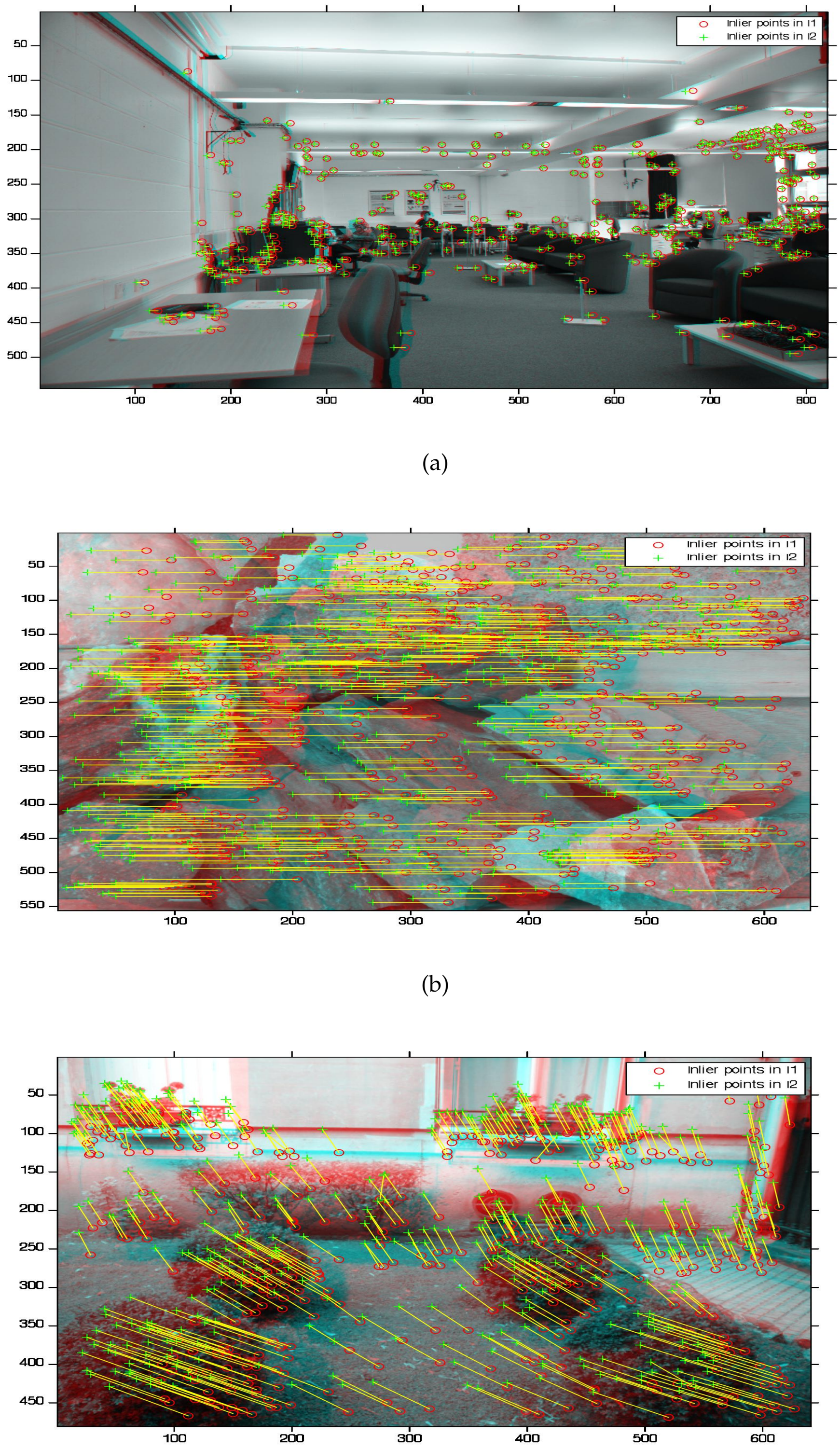
(c)

(d)

(e)

(f)

Figure 4.2: Sets of reference images. Each set has a left and right image presented along the same row.

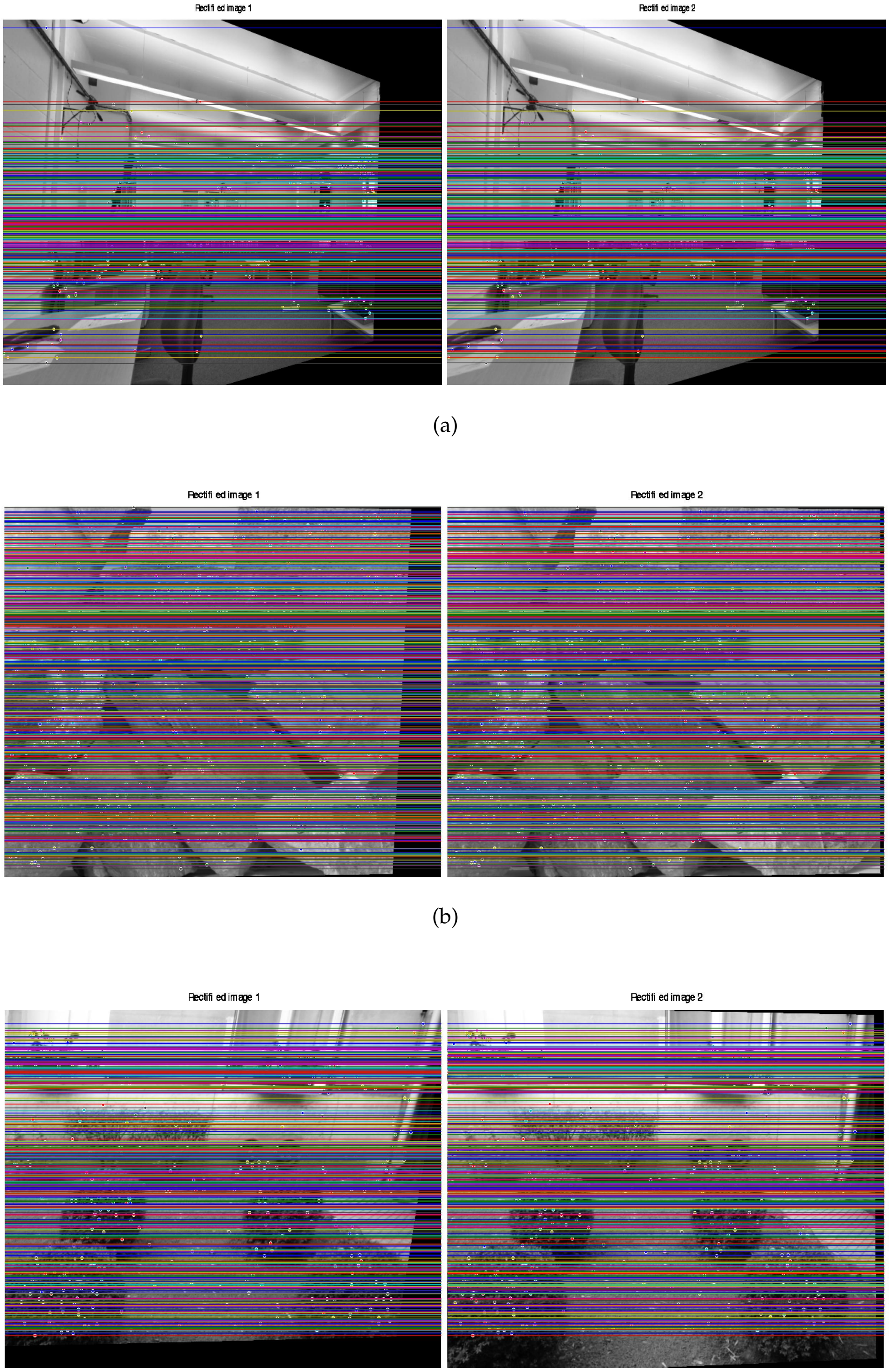


(a)

(b)

(c)

Figure 4.3: Composite image of each set in Figure 4.2 with corresponding inliers. The inliers are represented with a small red circle and a green and yellow plus sign for left and right images respectively.

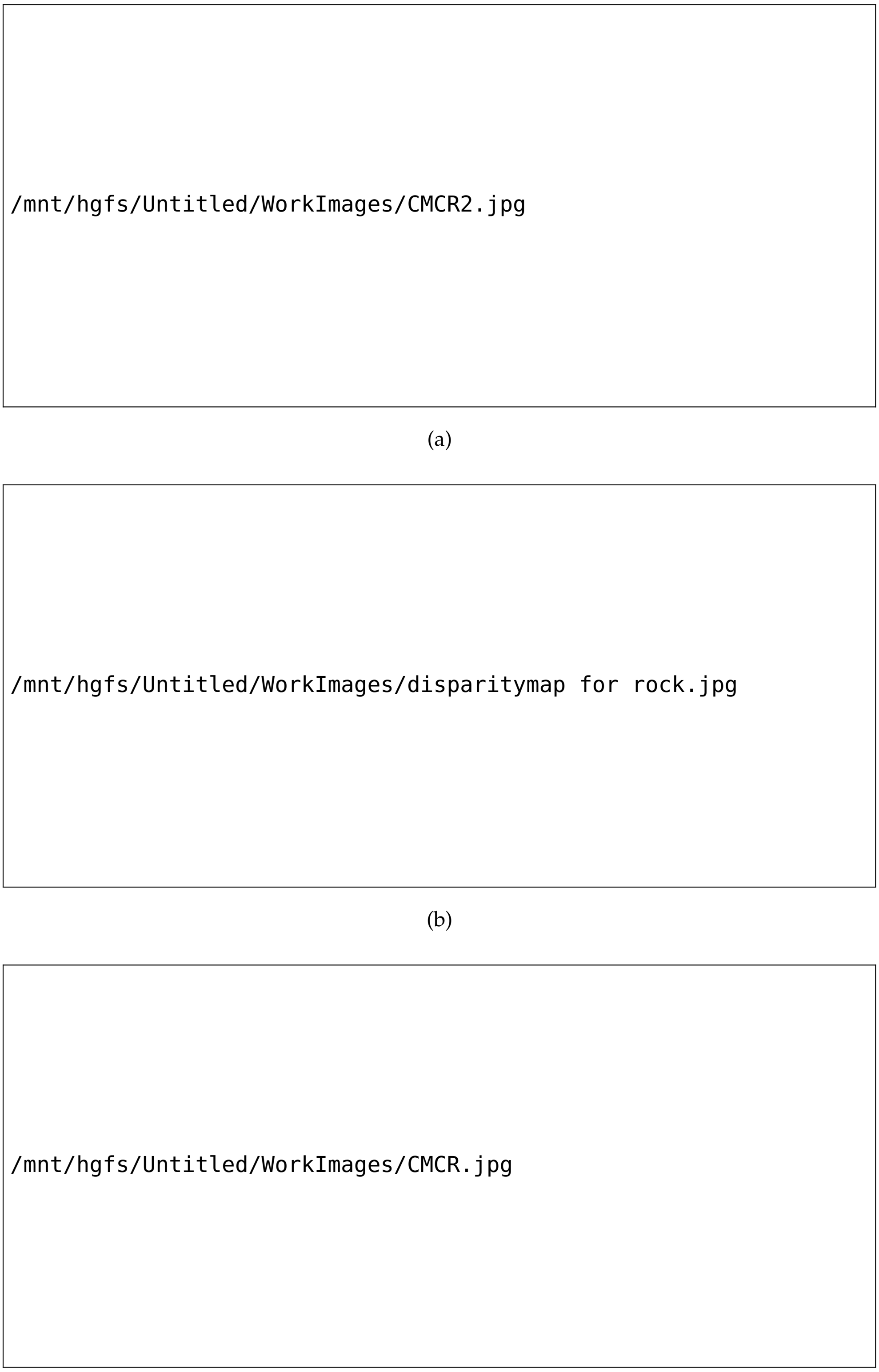


(a)

(b)

(c)

Figure 4.4: Rectified pair of each set in Figure 4.2 with corresponding inliers and epipolar lines. Any two corresponding inliers in the left and right images are at the same horizontal level and are joined using epipolar lines.

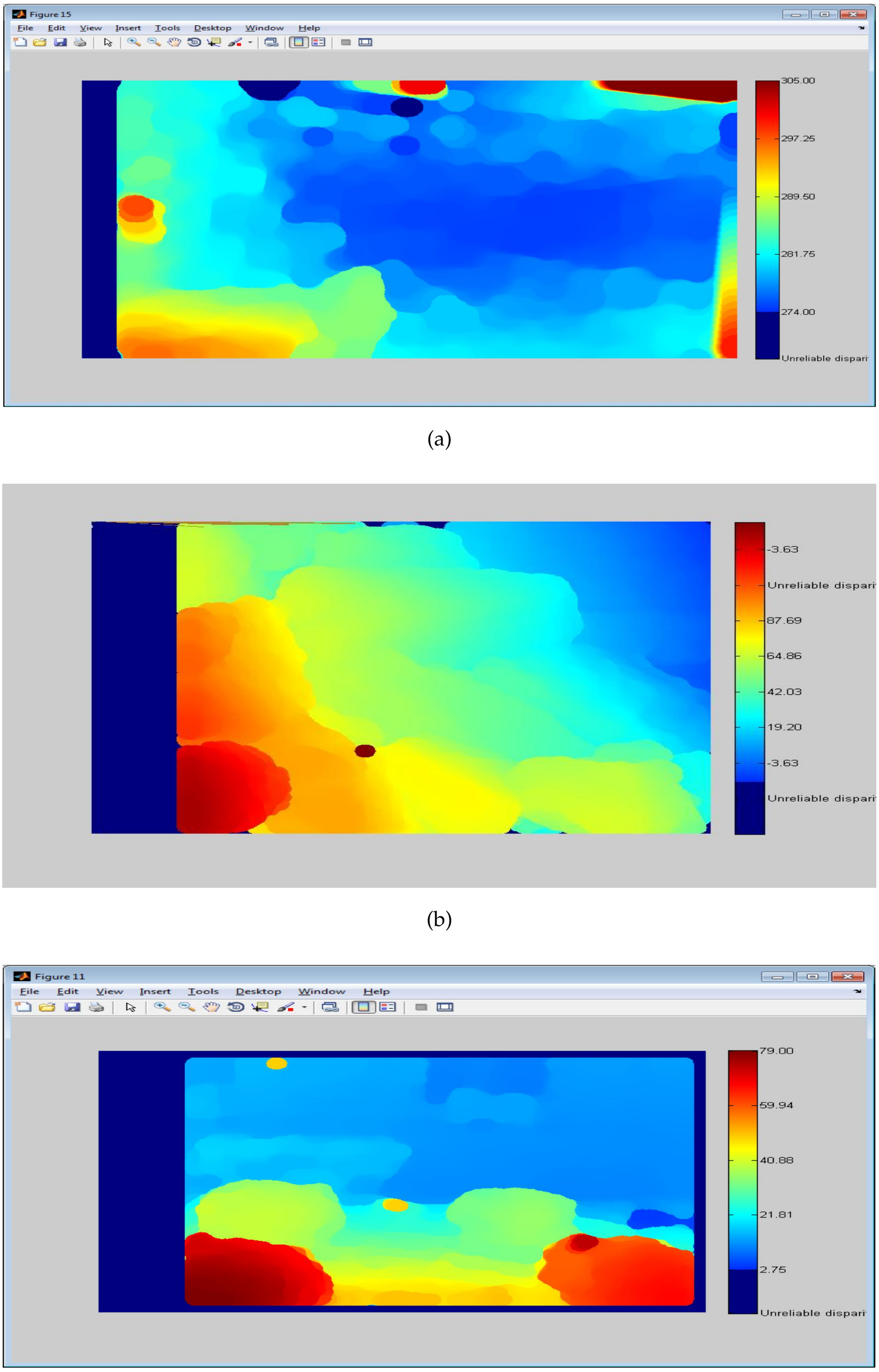

/mnt/hgfs/Untitled/WorkImages/CMCR2.jpg
/mnt/hgfs/Untitled/WorkImages/disparitymap for rock.jpg
/mnt/hgfs/Untitled/WorkImages/CMCR.jpg

(a)

(b)

(c)

Figure 4.5: Depth map of each set in Figure 4.2 using global method in which energy function minimisation is done via dynamic programming.

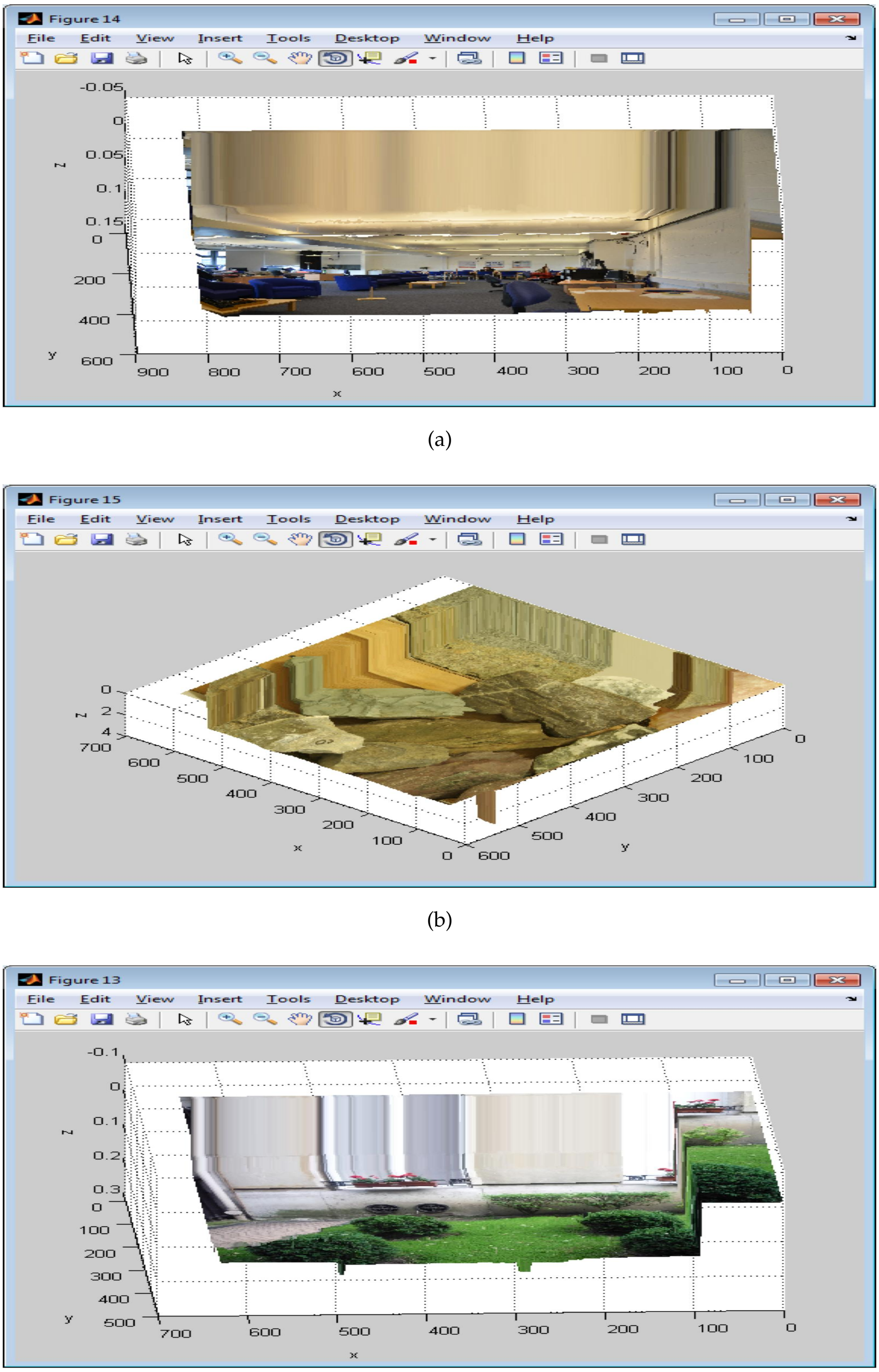


(a)

(b)

(c)

Figure 4.6: Depth map of each set in Figure 4.2. based on the method presented in this section.



(a)

(b)

(c)

Figure 4.7: Rendered images from the sets of images in Figure 4.2.

a vanishing point somewhere at the middle of the blue colour region. For (b) the depth increases from the bottom left hand corner of the figure (region in red colour) towards the upper right hand corner. Rendered images are presented in Figure 4.7.

In Figure 4.5, depth map is generated based on global method. The energy function is minimised based on dynamic programming. By comparison with the depth maps of Figure 4.6, generated based on the idea proposed in this work, the quality of the depth maps is low. Therefore, the depth maps constructed with reference to the work of this thesis is subjectively better in quality. Also, the speed of implementation of dynamic programming is longer with respect to depth map generation

4.2.4 *Conclusion*

This work has studied the effect of region match measures SAD, SSD, and NCC on the reliability and accuracy of image corresponding points which significantly affect the quality of rendered images. The combined use of both geometric and epipolar constraints helped to significantly reduce the number of outliers. Subjectively, NCC does better in reducing artifacts compared with SAD and SSD. In experiments, it is observed that the SSD method can yield a reasonably accurate depth map compared with SAD when the scene has a shallow depth. Also, the quality of the depth maps generated based on the method presented in this work is subjectively better than the results obtained by the use of dynamic programming approach. Also, there is a considerable reduction in implementation time in favour of the method presented in this work.

4.3 GENERATION OF THREE-DIMENSIONAL PANORAMIC IMAGE

4.3.1 *Introduction*

Panoramic image of scenes are widely common. What is not common is depth embedded panoramic image. Therefore, the generation of panoramic image characterised with depth perception is the primary focus of the discussion presented in this section. A panoramic view is a synthetic wide-angle camera realised mostly in software. Single and double cameras have been separately employed in the acquisition of image samples. In either case, the realisation of 3D panorama is achieved at two broad levels. First, two stereo panoramas are generated through image stitching by way of cylindrical projection as opposed spherical. In this stage, use is made of feature detection and extraction method presented in section 4.2. Other process performed in this stage are matching and compositing. Second, anaglyph is carried out in order to superimpose the generated stereo panoramic images on one another.

4.3.2 *3D Panoramic Image Implementation*

The implementation of this work is carried out at two broad stages as shown in **??**. First is the generation of a pair of panoramic views of the same scene. This stage comprises of using two Nikkon D7000 cameras (left and right) to acquire image samples which are subsequently written into a file. This is then followed by correction of radial distortion and projection onto a cylindrical surface. The correspondence problem between any pair of images is solved through feature detection, extraction, and matching [199, 295–300]. This leads to computation of homography and motion model. Finally left and right panoramic images are composed seamlessly.

The second stage at broad level is the anaglyph composition. Two colour panoramic views left (1) and right (2) are used to construct colour anaglyph.

Trimming adjustment is also used to vary the horizontal disparity until a comfortable and natural looking image is obtained. In the case of colour anaglyph, the RGB components are maintained even after the coding operation.

4.3.2.1 *Detailed Steps of Panoramic Image Generation*

- The first crucial step in the generation of any panoramic view is the acquisition of image samples of a scene through 360^o camera panning. The panning is done around an axis which is perfectly normal to the ground. Several images capture different portions of the same scene, with an over-

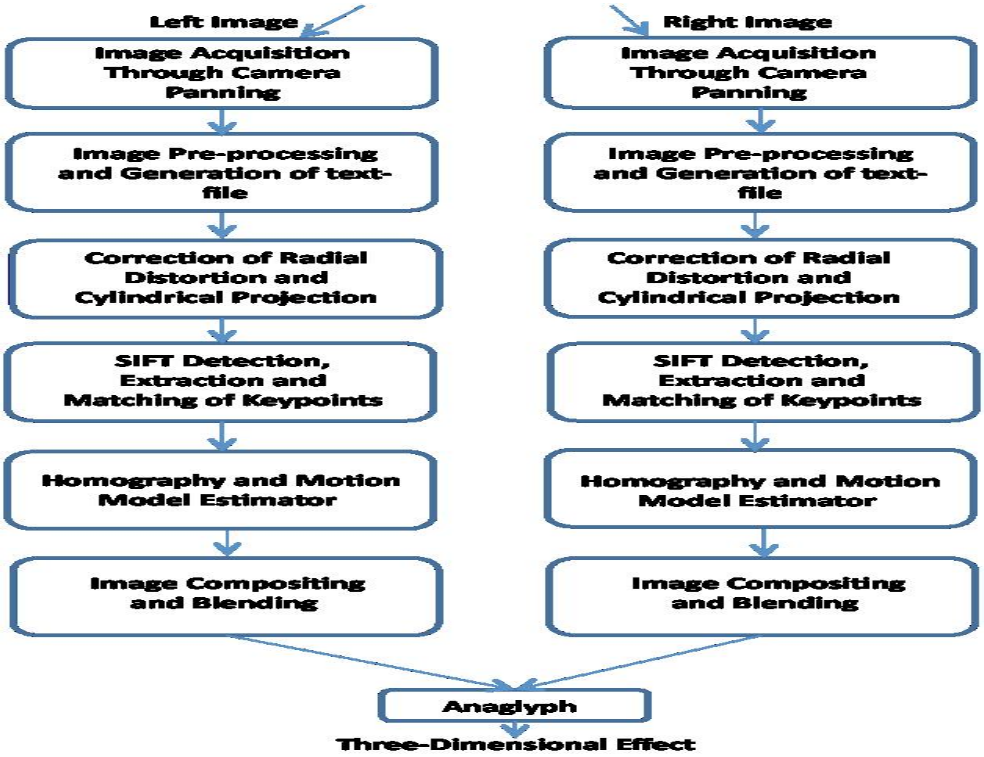

Left Image
Image Acquisition Through Camera Panning
Image Pre-processing and Generation of text-file
Correction of Radial Distortion and Cylindrical Projection
SIFT Detection, Extraction and Matching of Keypoints
Homography and Motion Model Estimator
Image Compositing and Blending
Right Image
Image Acquisition Through Camera Panning
Image Pre-processing and Generation of text-file
Correction of Radial Distortion and Cylindrical Projection
SIFT Detection, Extraction and Matching of Keypoints
Homography and Motion Model Estimator
Image Compositing and Blending
Anaglyph
Three-Dimensional Effect

Figure 4.8: Block diagram for generating 3D effect from two panoramic views.

lap region viewed in both images. A path description of each image location is then contained in a text file.

- In this work, a cylinder is used as the projection surface. This allows for an 180^o by 360^o field of view enhancement. This step is then followed by correction of radial distortion associated with image. Two types of radial distortion can be corrected: Barrel and pincushion. In "barrel distortion", image magnification decreases with distance from the optical axis. The apparent effect is that of an image which has been mapped around a sphere (or barrel). In pincushion distortion, image magnification increases with the distance from the optical axis. The visible effect is that lines that do not go through the centre of the image are bowed inwards, towards the centre of the image, like a pincushion. Brown's (1972) extension of Magill's formulation for variation of radial distortion with focusing still remains potentially attractive. This is in spite of the re-verification by [301–303], with data of much higher precision than the previous investigations. Brown's distortion model is generally used to correct radial distortion. This is expressed in (4.9). (x_d, y_d) describes the coordinates of the distorted image while(x_u, y_u) is for undistorted.

$$\left\{ \begin{aligned} x_d &= x_u\left(1 + \left(k_1 r^2\right) + \left(k_2 r^4\right)\right) \\ y_d &= y_u\left(1 + \left(k_1 r^2\right) + \left(k_2 r^4\right)\right) \end{aligned} \right\} \tag{4.9}$$

- Fundamentally, image registration involves the establishment of a motion model which allows for proper integration of useful information from multiple images of the same scene taken at different times, from different viewpoints and/ or by different sensors. Depending on the area of application, image registration can be either multi-temporal analysis (different time), multi-view analysis (different viewpoints), scene to model registration (images of a scene and its model are registered), and multi-modal analysis (different sensors are used in image acquisition).

- It is well established in literature that irrespective of application area, an image registration is usually implemented under four steps namely: feature detection in which the descriptive image regions called feature points are detected, feature matching, motion model estimation, and image resampling and transformation. The published work of [199] has the established reputation of detecting and using a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors. It is characterized by detection and localization of keypoints in different scale space images, followed by the assignment of an orientation to each keypoint using local image gradients. Then a keypoint descriptor is assembled from the local gradient values around each keypoint using orientation histograms. SIFT, [199], has been used in this work for feature detection and extraction.

- The accuracy of panoramic mosaic to a large extent is dependent on the image matching technique employed in the establishment of correspondence between one or several images and a reference. Correspondence between any two images is established using feature-based matching. The features common under geometric constraints to the two images called inliers serve as a prerequisite for the computation of projective matrix and subsequently the motion model. The inliers are computed using an algorithm for robust fitting of models in the presence of many data outliers [293]. The composition of the motion model is such that it allows an image to be transformed with respect to another which is considered to be the reference. The obvious consequence of this process is that the transformed image can then be stitched to the reference image at the proper coordinate points.

- Image composition is the stitching of the transformed image obtained through motion model to the reference image at the computed coordinate. This is implemented using the proposed continuous division method.

- Image blending is implemented using image feathering technique extensively discussed in [304]. This is done to allow for a smooth transition from one image to the other across the transition boundary. In Feathering or alpha blending which is expressed in (4.10), the mosaic image is a weighted combination of the input images I_1 and I_2. The weighting coefficients w, vary as a function of the distance from the seam [204]. Pyramid blending involves the combination of different frequency bands with different alpha masks. Lower frequencies are mixed over a wide region, and fine details are mixed in a narrow region. This produces gradual transition in lower frequencies, while reducing edge duplications in textured regions

$$I\left(i,j\right) = \left(1 - w\right) I_1\left(i,j\right) + wI_2\left(i,j\right) \tag{4.10}$$

4.3.2.2 *Camera Focal Length and Distortion Parameters*

- Camera focal length used in image stitching algorithm is usually expressed in pixels. This can be obtained using the relationship in (4.11) where focal length in pixels and mm are denoted by $f_{c[pix]}$ and $f_{c[mm]}$ respectively. $N_{[pix]}$ is the camera sensor dimension in pixels in either x or y direction

$$f_{c[pix]} = \frac{f_{c[mm]}\left(N_{[pix]}\right)}{size\,of\,CCD} \tag{4.11}$$

- Camera calibration aims for a determination of the transformation parameters between the camera lens and the image plane as well as between the camera and the scene based on the acquisition of images of a calibration rig with a known spatial structure. Outlined in [305, 306], Tsai, Zhang, and Bouguet techniques are especially suited for fast and reliable calibration of standard video cameras and lenses which are commonly used in computer vision applications. Camera calibration toolbox for Matlab has been used to confirm the focal length and coefficient parameters of the lens used in the acquisition of image samples.

- The camera focal length in mm and lens distortion coefficients k_1 and k_2 are obtained through calibration process.

4.3.3 *Experimental Results and Discussion*

4.3.3.1 *Radial Distortion and Full Homography*

Each image sample has a resolution of 2464×1632. Brown's distortion model is used for the correction of radial distortion with a focal length of 385.153pixels, and distortion coefficient parameters $k_1 = -0.023$, $k_2 = 0$.

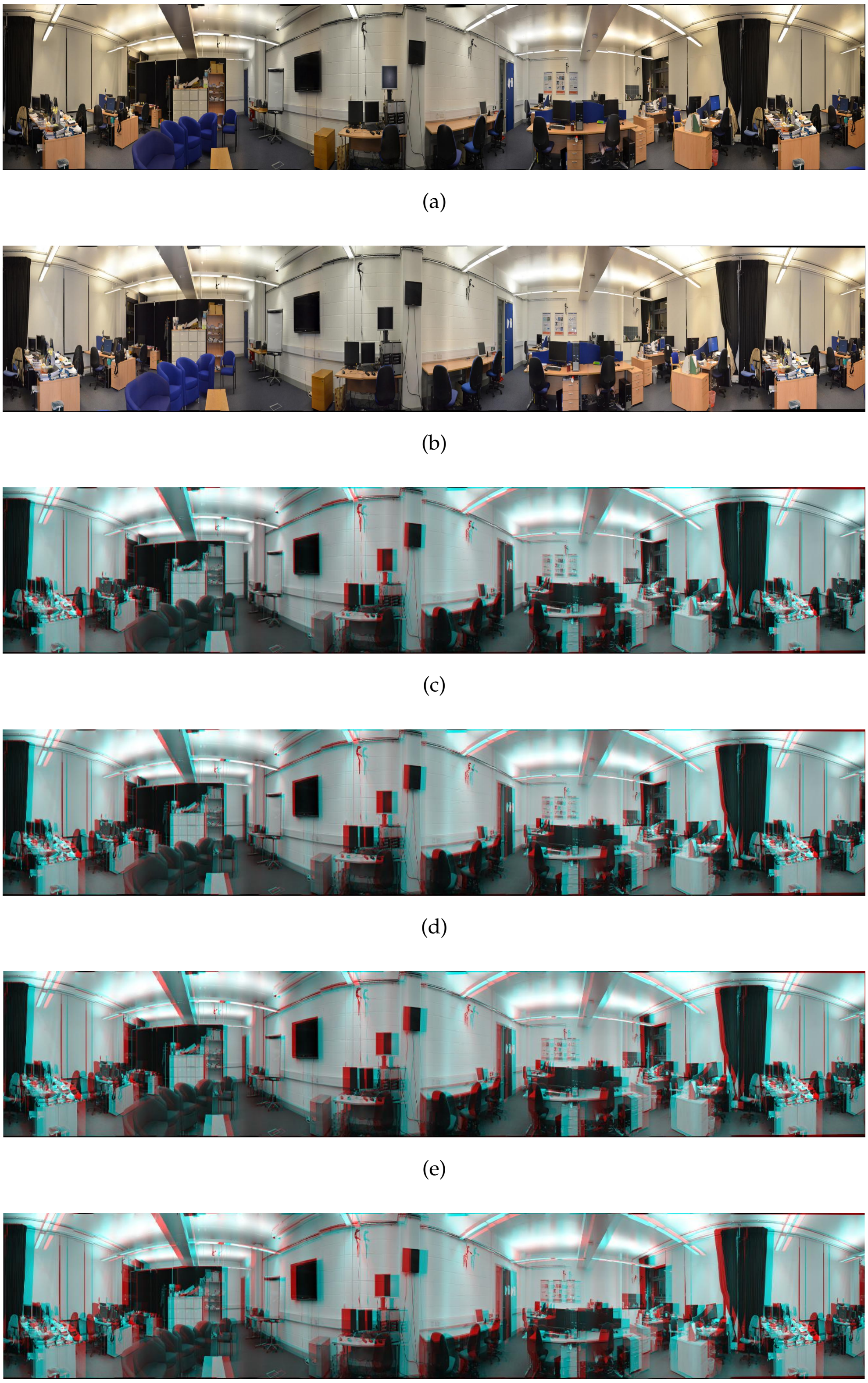
Feature descriptors are extracted using the SIFT algorithm. Candidate correspondence points between any two images are computed based on these descriptors. The candidacy of a correspondence point is confirmed if the ratio of the Euclidian distance of the top two nearest neighbours is less than a threshold of 0.6. Then output of RANSAC algorithm is used to compute full homography.

The results and appropriate discussion about this work are presented as follows. Two sets of image samples at a resolution of 2464×1632 are acquired using two compact digital cameras (Nikkon D7000) mounted on a single tripod. Each set contains thirty six images. Image acquisition was made with stereoscopic distance of $130mm$, $140mm$, $150mm$, and $160mm$.

Thirty six sample images have been used for each of the constructed panoramic views with the first and last images being repeated. One important observation is that, in the panoramic views of Figure 4.9 through to Figure 4.11, the observed perspective projection is different for similar objects of the same physical size located at almost the same point in the scene. This means that (a) and (b) parts of Figure 4.9 through to Figure 4.11 are different perspective panoramic views of the same scene.

Also the binocular depth which is observed through the use of a pair of anaglyph glasses increases from (c) through to (f) in the just mentioned figures. What can be easily noticed is the pop out effect from the screen.

Another set of experiments in simulation is performed to generate depth perception from a panoramic image constructed from a set of images captured using a single panning camera. For each image set, one real panoramic image is generated. From the real panorama, a virtual one with disparity is further generated. These are then superimposed on each other to create 3D effect. Some of the obtained results are shown in Figure 4.12 through to Figure 4.17. The figures contain two out-door and one indoor scenes. Each scene is demonstrated in several depth perceptions.



(a)

(b)

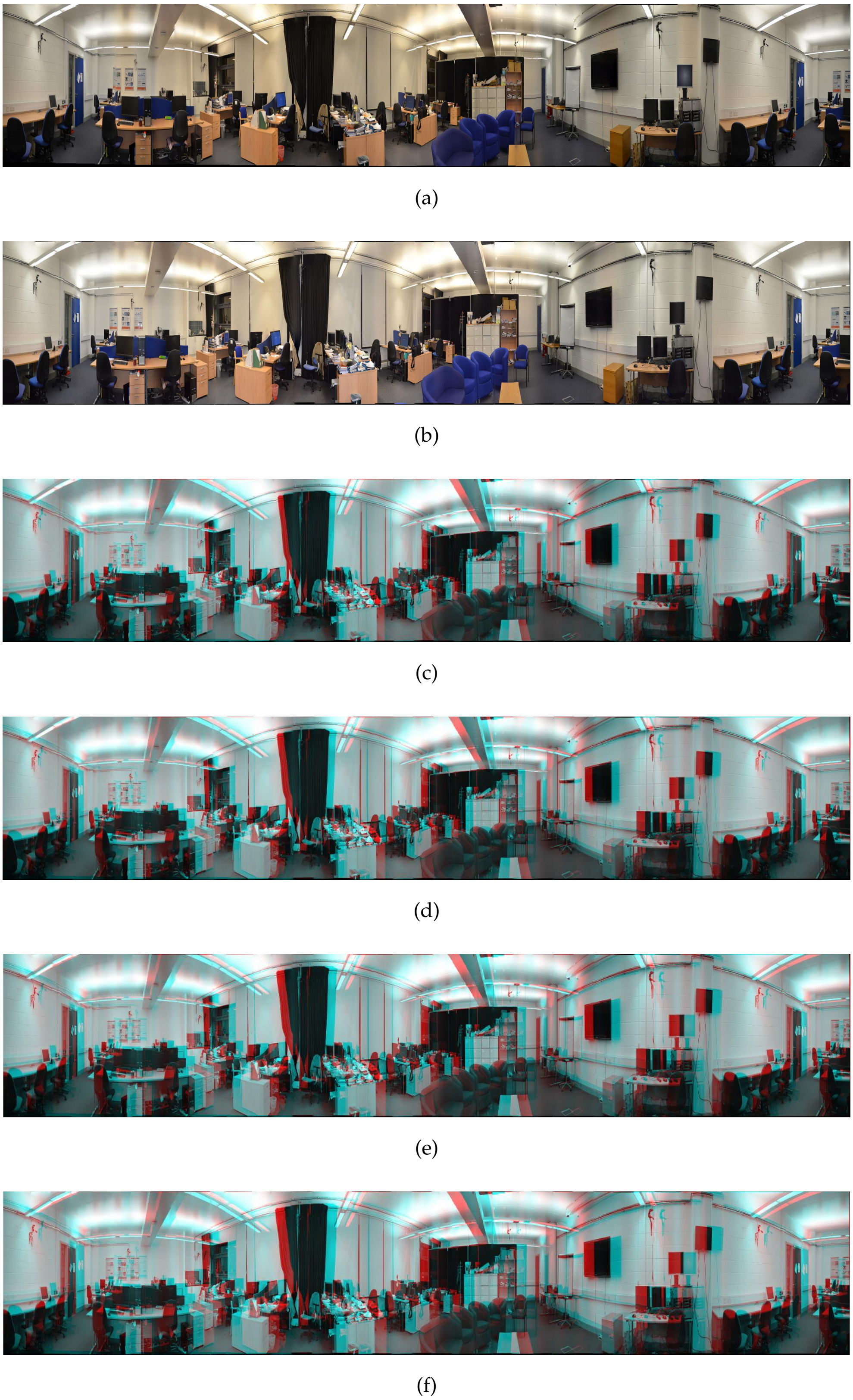
(c)

(d)

(e)

(f)

Figure 4.9: Binocular depth due to stereo panoramic image. (a) and (b) are left and right images. (c), (d), (e), and (f) are images with different 3D depths.



(a)

(b)

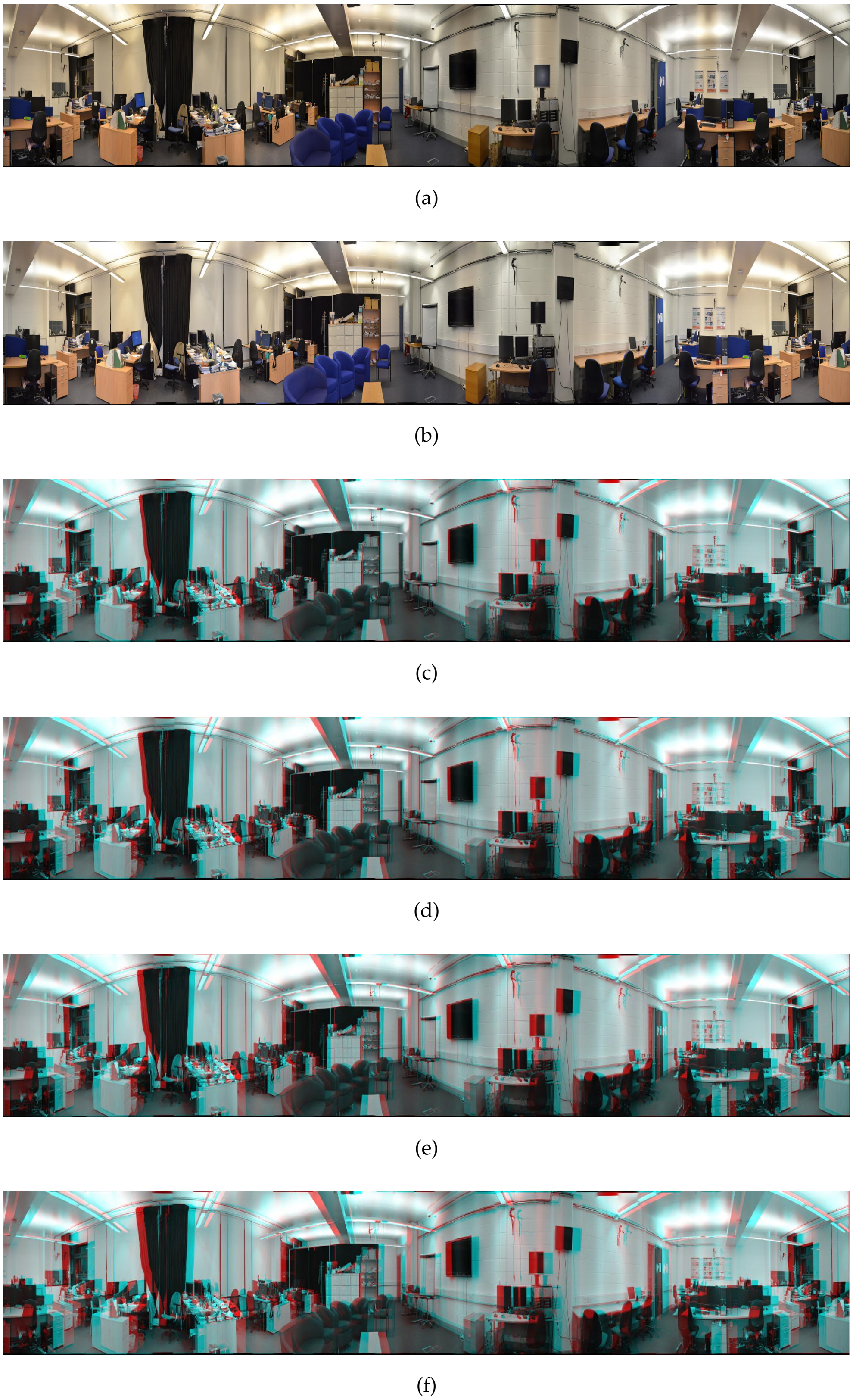
(c)

(d)

(e)

(f)

Figure 4.10: Binocular depth due to stereo panoramic image. (a) and (b) represent left and right images. (c), (d), (e), and (f) are images with different 3D depths.



(a)

(b)

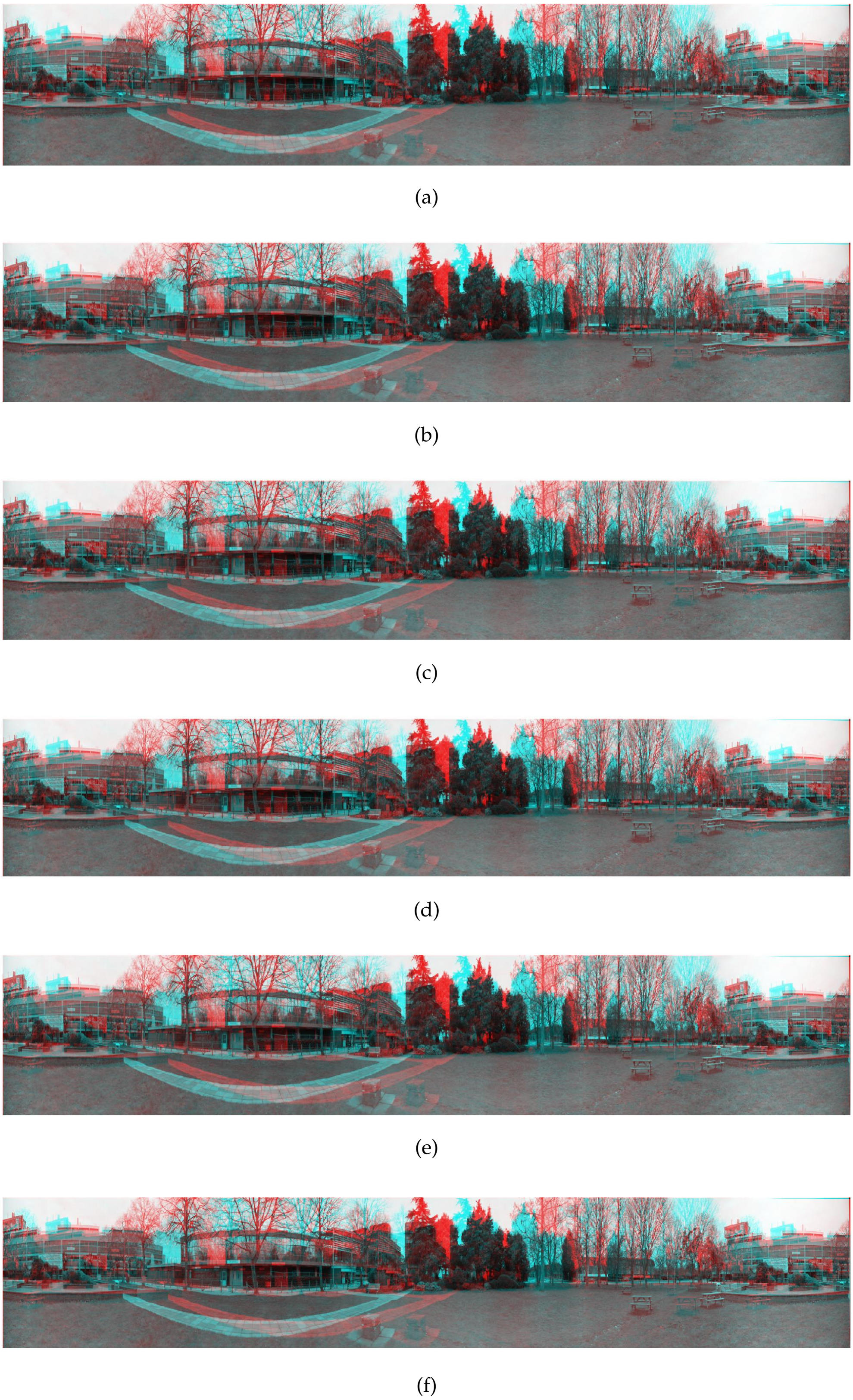
(c)

(d)

(e)

(f)

Figure 4.11: Binocular depth due to stereo panoramic image. (a) and (b) are left and right images. (c), (d), (e), and (f) are images with different 3D depths.



(a)

(b)

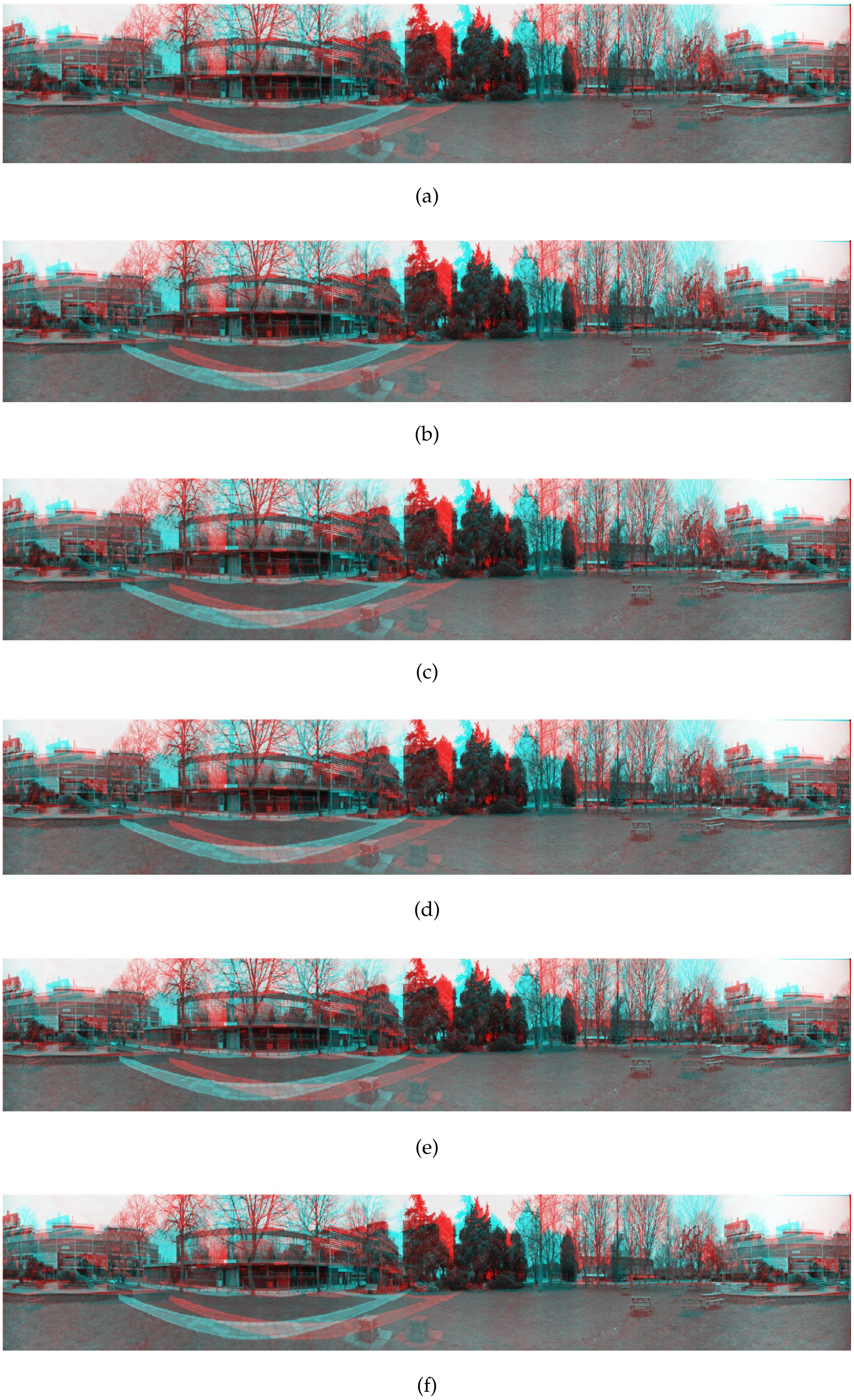
(c)

(d)

(e)

(f)

Figure 4.12: 3D effect from a real and virtual panoramic images constructed from multiple image samples obtained from a single camera. The depth perception increases from (a) through (f).



(a)

(b)

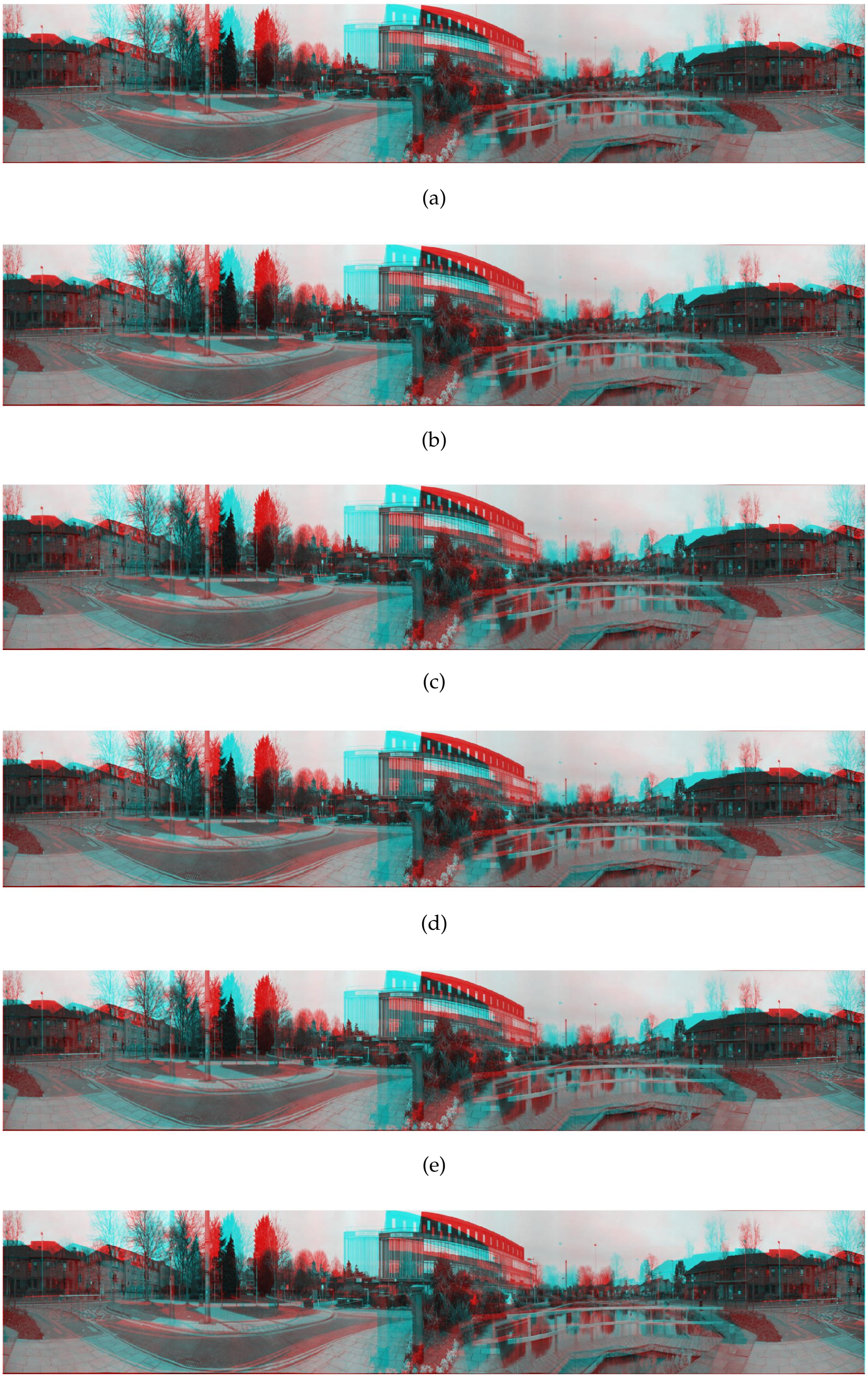
(c)

(d)

(e)

(f)

Figure 4.13: 3D effect from a real and virtual panoramic images constructed from multiple image samples from a single camera. The depth perception increases from (a) through (f).



(a)

(b)

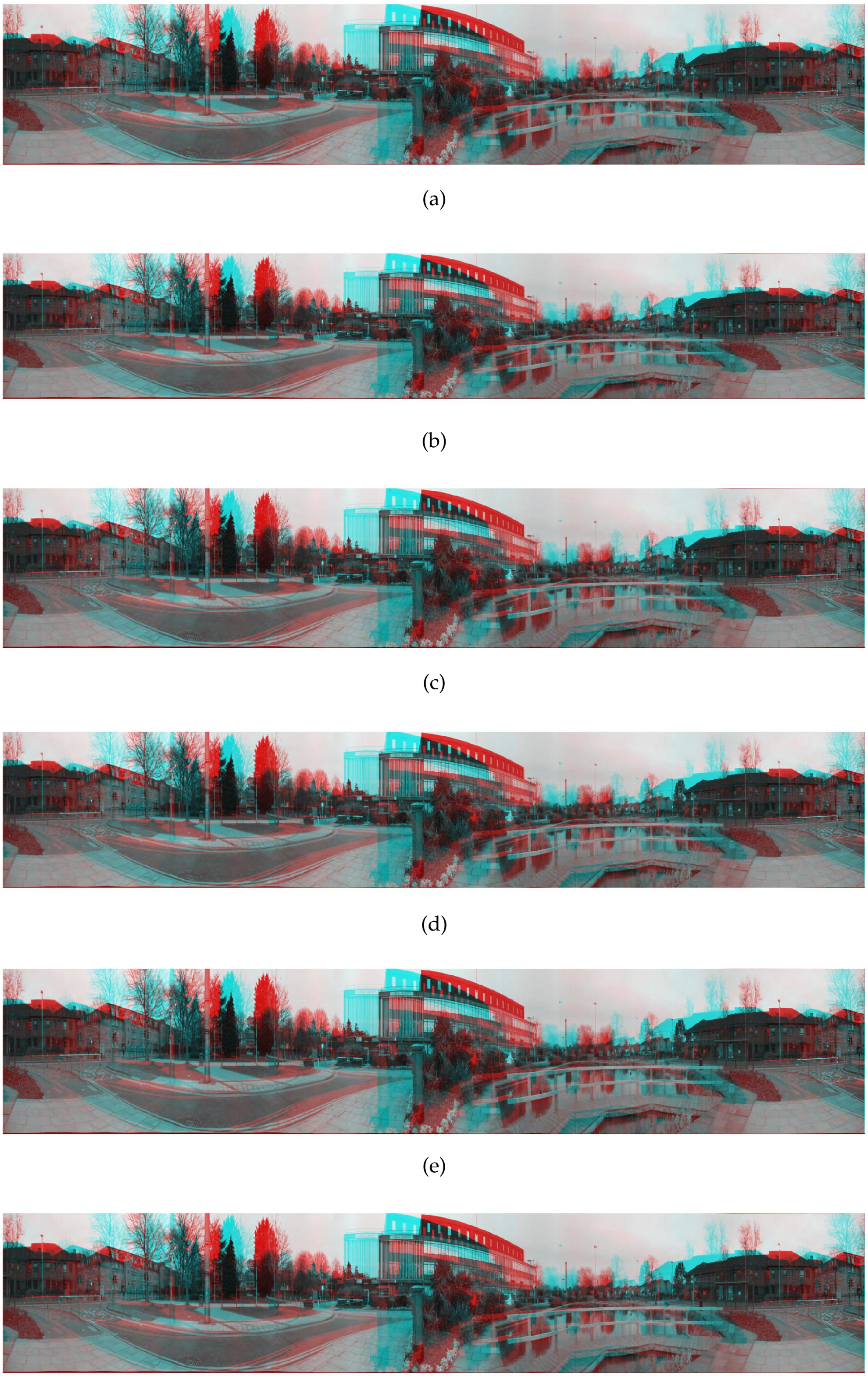
(c)

(d)

(e)

(f)

Figure 4.14: 3D effect from a real and virtual panoramic images constructed from multiple image samples from a single camera. The depth perception increases from (a) through (f).



(a)

(b)

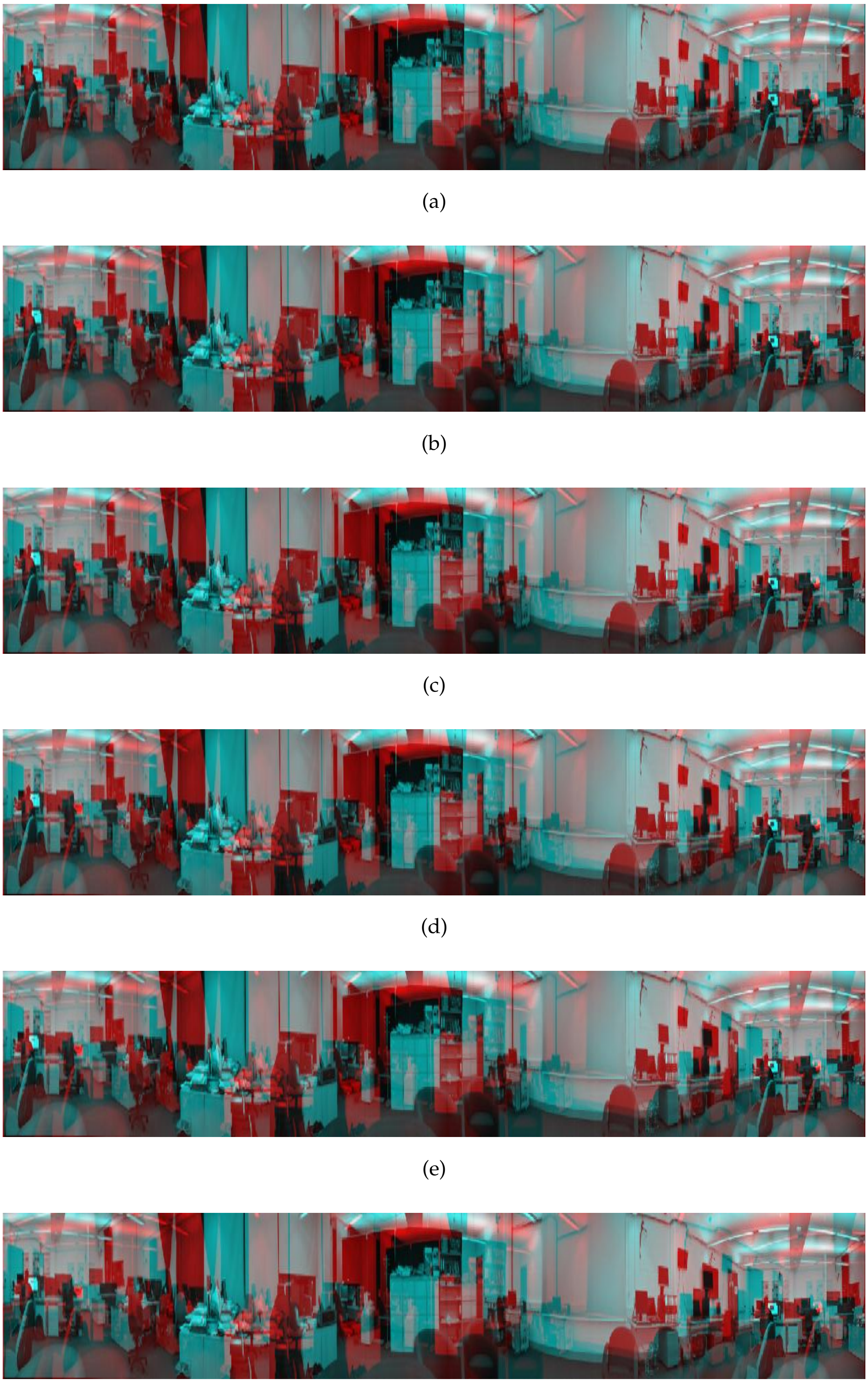
(c)

(d)

(e)

(f)

Figure 4.15: 3D effect from a real and virtual panoramic images constructed from multiple image samples from a single camera. The depth perception increases from (a) through (f).



(a)

(b)

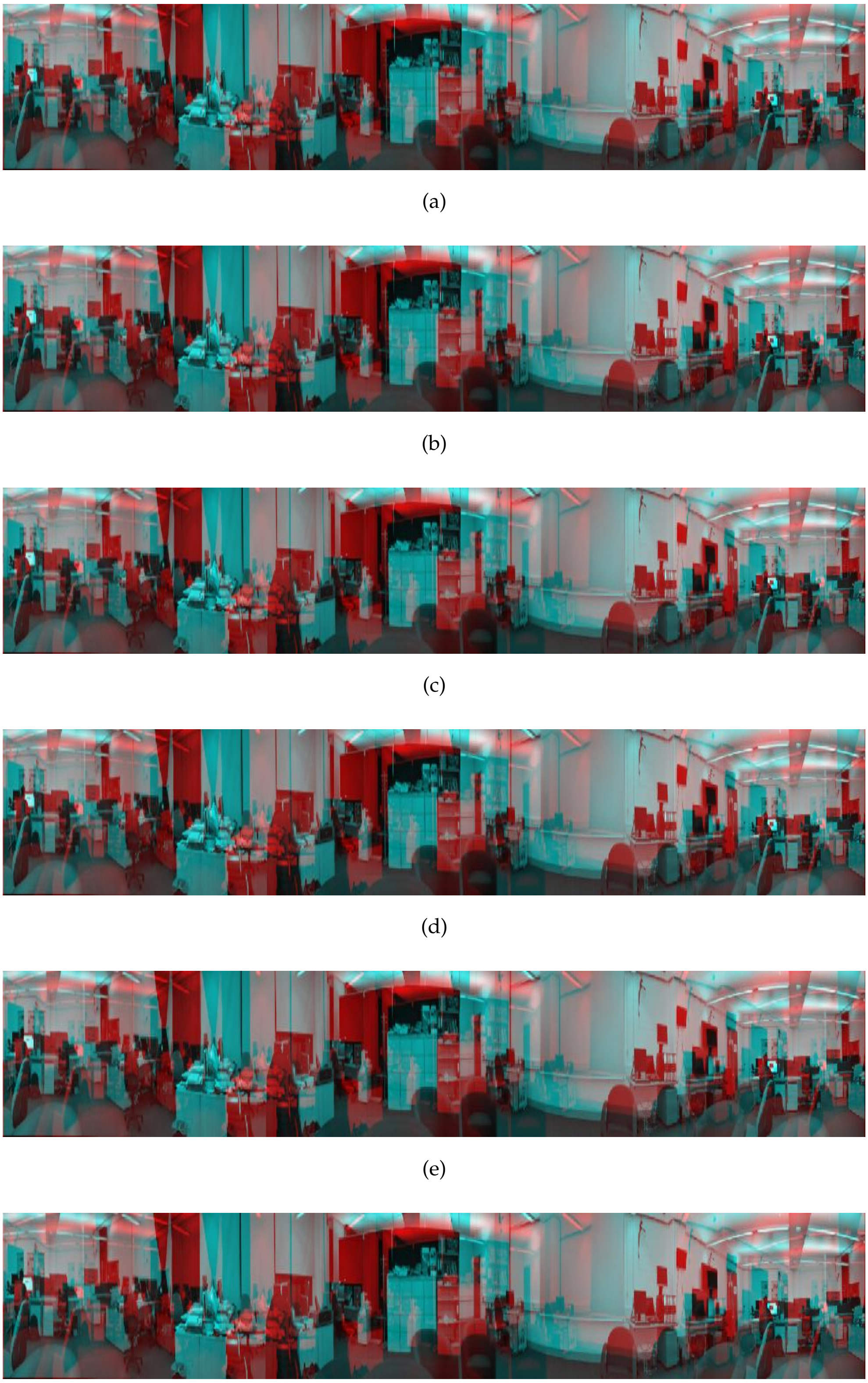
(c)

(d)

(e)

(f)

Figure 4.16: 3D effect from a real and virtual panoramic images constructed from multiple image samples from a single camera. The depth perception increases from (a) through (f).



(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.17: 3D effect from a real and virtual panoramic images constructed from multiple image samples from a single camera. The depth perception increases from (a) through (f).

4.3.4 *Objective and Subjective Image Quality Assessment*

A commonly used objective measure of image quality is modulation transfer function MTF. The definition and description of this parameter have been presented in the appendix section of this thesis.

Hamerly's edge raggedness, or tangential edge profile; Granger and Cupery's Subjective Quality Factor (SQF) derived from the second moment of the line spread function; and SQF derived from Gur and O'Donnell's reflectance transfer function are the perceived image quality metric which stand-out in literature. The subjective assessment method embedded in Imatest software is based on SQF presented in [307]. The software also has a provision for performing SQF assessment on either a part of, or the whole image by taken into consideration the combined effects of MTF, the contrast sensitivity function of HVP, the height dimension of the image and the viewing distance. This idea stems from [307] and has been evaluated in [308] where it is explained that a graphics target and text in a variety of fonts was viewed by a panel of eight judges who are untrained. An integer score was assigned to each print based on its overall quality. Analysis of the metrics revealed that SQF by [307] had the highest correlation with panel rank, and achieved a level of precision approaching single-judge error, that is, the ranking error made by an individual judge.

An image with a resolution (dimension) of 1666×449 pixels ($W \times H$) is used in the results presented in Figure 4.18 where the region of interest is 1666×246 pixels. The Imatest software computes the viewing distance most suitable for subjective assessment of the image. The text on the lower right-hand corner of the larger graph contains other calculation details and image properties. The image is viewed from a distance based on a base distance of $30cm$ determined by the Imatest software. The larger graph on the left contains a plot of SQF and Viewing distance against picture (image) height.

For the SQF versus picture height plot, two edge response curves constitute the SQF information shown. The plot in black is for SQF without standardised sharpening and the one in red is for SQF with sharpening. Basically, sharpening is meant for increasing or decreasing the contrast of an image at boundaries by reducing the rise distance. The reason for the use of standardised sharpening is to eliminate any strong transient that may be associated with the change in SQF as a result of the change in image height. Therefore, sharpening is brought under control. Standardised sharpening can cause a decrease in sharpness and hence leads to a decrease in SQF as can be seen in the graph. Based on the partial region of interest considered, standardised sharpening is achieved using pixel shift of 130. This is referred to as sharpening radius. It is observed that sharpening has caused a decrease of 2% in the SQF from an average value of 8%. In the case of Viewing distance versus picture height plot, from the light blue curve, a typical impression of perceived sharpness can be observed since the Viewing distance is proportional to the square root of the picture height. The plot on the upper right represents MTF values at different spatial frequencies, without and with standardised sharpening.

In Figure 4.19, the region of interest considered is the entire image size of 1666×449. Using a base distance of $30cm$, the Viewing distance is computed. it is noticed that the decrease in SQF without and with the use of standardised sharpening is about 0.05%, however, the average SQF is approximately 1%. This is due to the large size of the sharpness radius used. Also the value of the MTF without and with standardised sharpening is almost the same.

For the 3D panoramic images shown in Figure 4.12 through to Figure 4.17, two samples of the subjective assessment are shown in Figure 4.20 and Figure 4.21 corresponding to partial and entire region of interest respectively. The interpretation of the graph is similar to the earlier explanation. The SQF without and with standardised sharpening is low and has almost the same value. There is also a considerable similarity in the MTF values at different spatial frequencies.

4.3.5 *Conclusion*

The generation of 3D effect from two panoramic views whose image samples are obtained from single and two synchronised cameras, has been demonstrated. Acquisition of image samples is carried out in both indoor and outdoor environ-

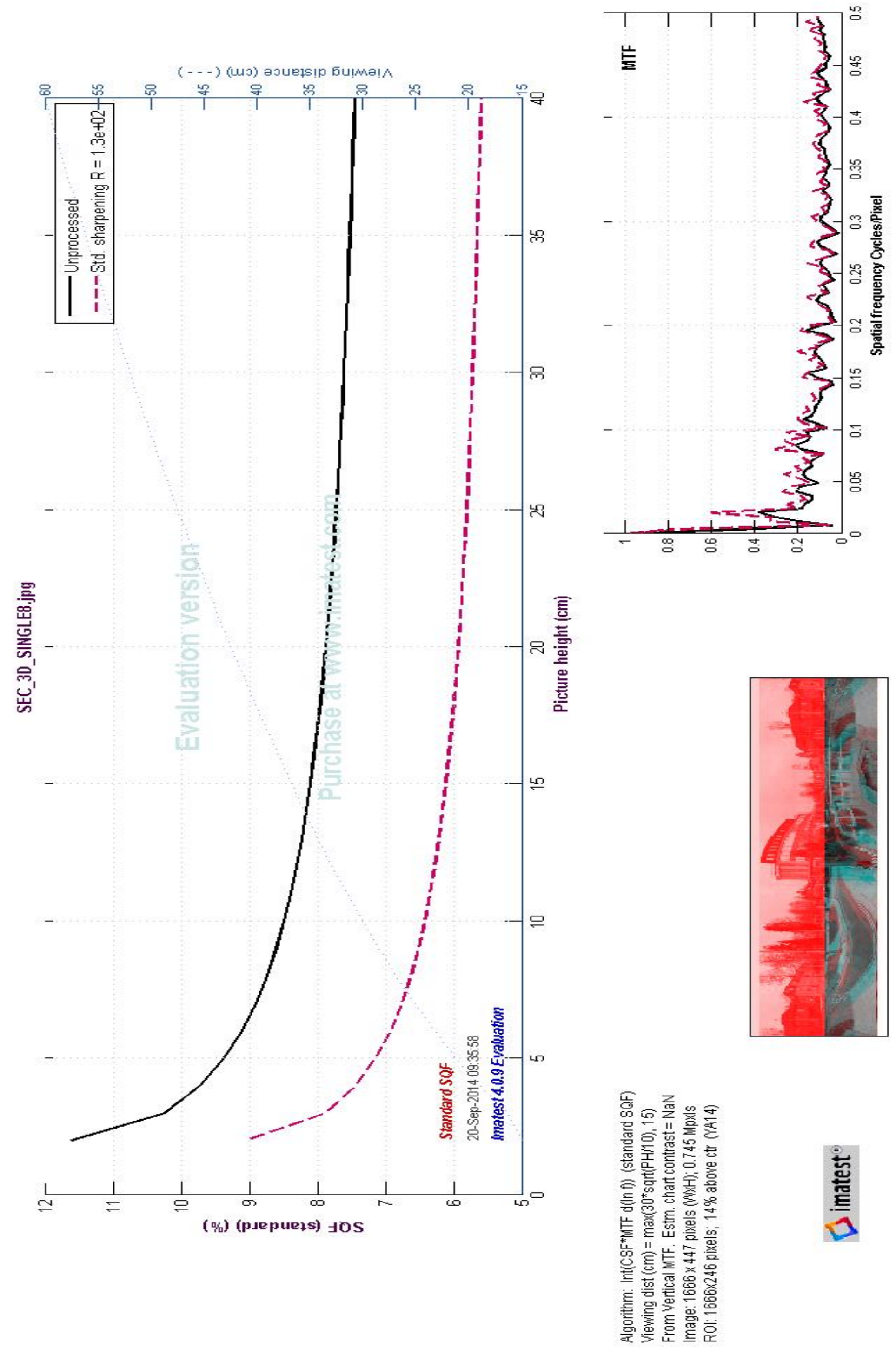


Figure 4.18: Subjective quality factor test for a 3D panoramic image whose image samples were obtained from two cameras simultaneously mounted on the same tripod (ganged or synchronised). The region of interest used in this analysis is determined by the Imatest software and is not the whole image.

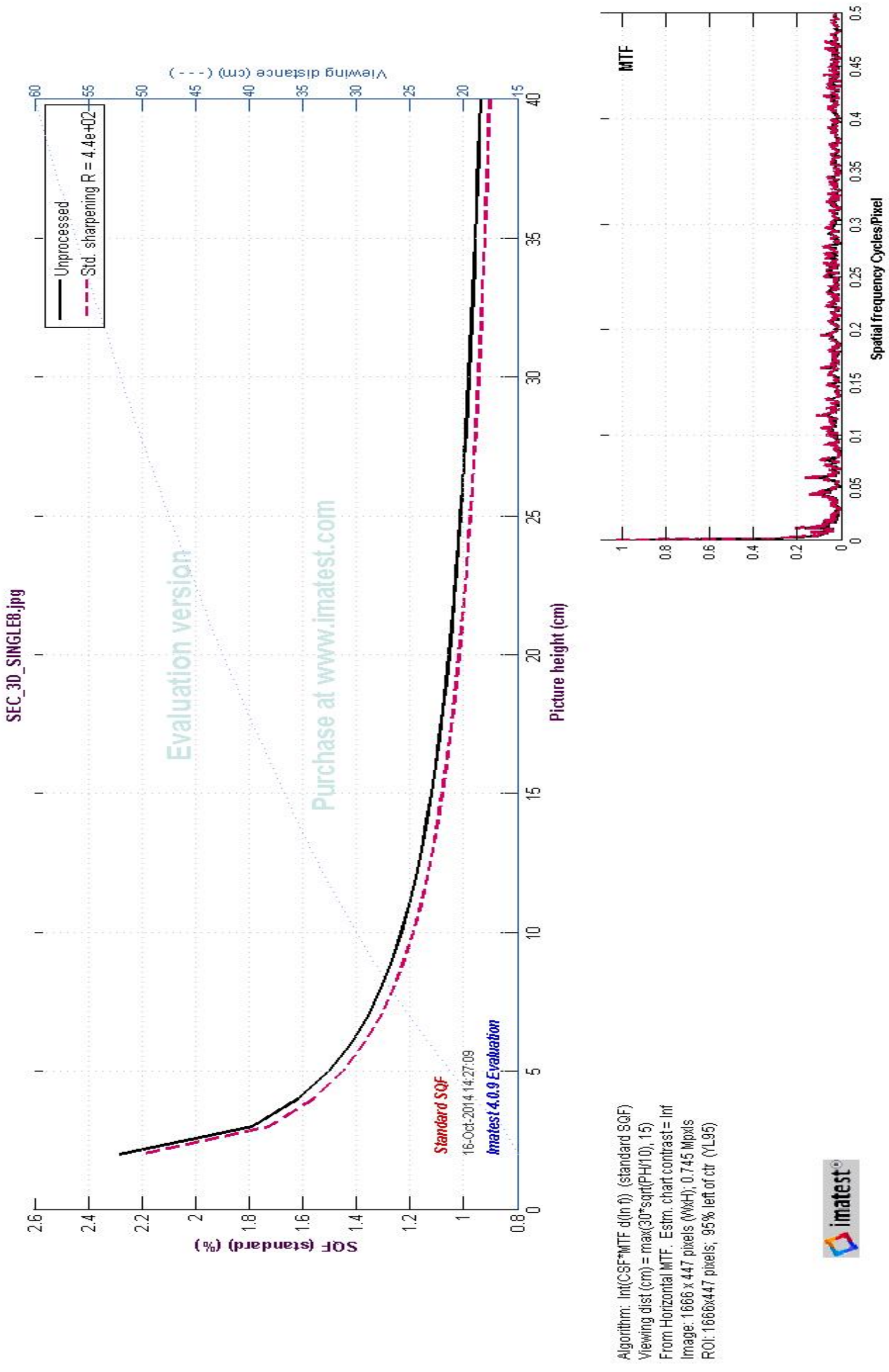


Figure 4.19: Subjective quality factor test for a 3D panoramic image whose image samples were obtained from two cameras simultaneously mounted on the same tripod (ganged or synchronised). The region of interest used in this analysis is determined by the Imatest software and is the whole image.

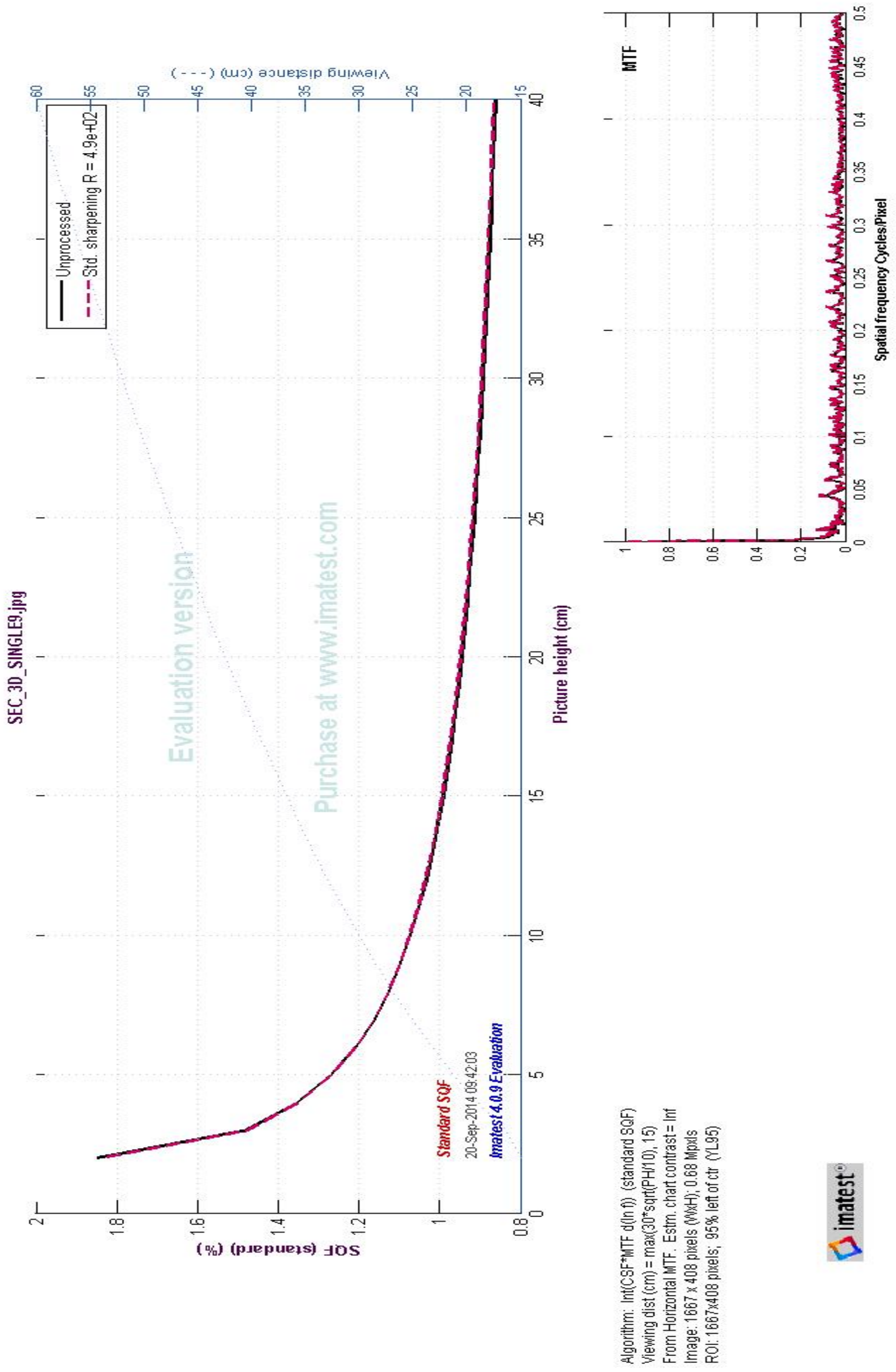


Figure 4.20: Subjective quality factor test for a 3D panoramic image whose image samples were obtained from one camera mounted on the same tripod. The region of interest used in this analysis is determined by the Imatest software and is not the whole image.

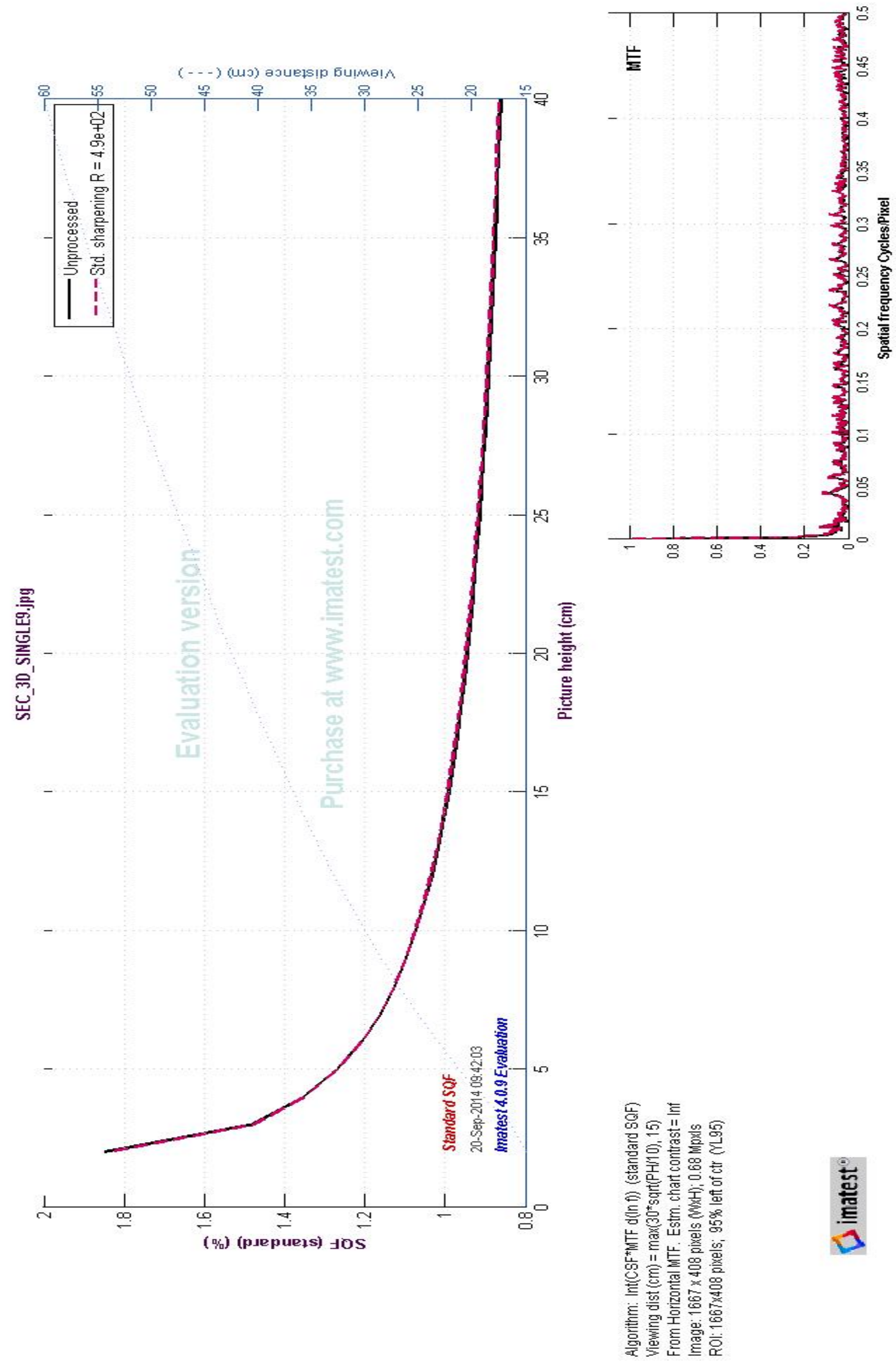


Figure 4.21: Subjective quality factor test for a 3D panoramic image whose image samples were obtained from one camera mounted on the same tripod. The region of interest used in this analysis is determined by the imtest software and is the whole image.

ments. Little or considerable vegetation has been used as a criteria for the choice of outdoor environment in this work. Prior to the anaglyph composition, each of the two panoramas is obtained through image stitching. After the multiple image acquisition stage, the stitching process starts with image preprocessing and text file generation. This is then followed by the correction of radial distortion and cylindrical projection. Further more, the detection, extraction, and match of interest points are performed on any possible pair of images. Estimation of homography and motion model are carried. The final stage of image stitching process is the image composition and blending. Therefore, all the mentioned steps have been implemented in the formation of each panoramic image which constitute a stereo panorama (left and right panorama). After the composition of a pair of panorama, anaglyph formation which will result in the perception of depth in the resultant panoramic image was performed. It is important to note that the quality of 3D effect largely depends on how well the multiple image samples have stitched. The subjective assessment based on SQF and obtained from Imatest software has provided considerable information regarding the perceived image quality. Also, based on a practical viewing experience, a fairly confident conclusion is that the resultant depth quality is good and does not cause much eye strain. It is important to mention that the image stitching method implemented in this section will be used in section 4.4.

4.4 COMPUTATION OF VIRTUAL ENVIRONMENT FROM STEREO-PANORAMIC IMAGE

4.4.1 *Introduction*

This work proposes depth image-based synthesis of virtual environment starting with stereo-panoramic views of a scene. It involves the use of two methods at the extremes of the IBR spectrum. These are IBR without and with geometry. A stereo-panoramic view is first generated by mosaicking a set of images containing both the optical and geometric properties of different parts of the scene. This is done using the image stitching methods presented in section 4.3. The second step which is centred on the second extreme of IBR is performed. In this step, the construction of a panoramic depth map carried out. A depth map of the scene contained in the stereo-panoramic view is constructed based on the normalisation of matching points and the singular value decomposition of fundamental matrix. Finally, the depth map is textured with one of the panoramic views to obtain a synthesized view. While a good quality depth map can only be generated if and only if the epipolar constraint condition for planar rectification is satisfied, it is also the determining factor for the computation of virtual environment from a stereo-panoramic view.

4.4.2 *Rectification and Epipolar Constraint*

he width component of the resolution of the stereo-panoramic image poses a serious challenge to the rectification process and hence satisfying the epipolar constraint becomes a problem. The simple and efficient planar rectification method is observed to be less efficient because of the rotational camera motion that is involved in panoramic views.

4.4.2.1 *Rectification and Epipolar Constraint*

Rectification remains a necessary step of stereoscopic analysis. It is a process which extracts epipolar lines and realigns them horizontally into a new rectified image. This allows subsequent stereoscopic analysis algorithms to easily take advantage of the epipolar constraint and reduce the search space to one dimension, along the horizontal rows of the rectified images. As critically observed in [309], the set of matching epipolar lines varies considerably and extracting those lines for the purpose of depth estimation can be quite difficult. The difficulty does not reside in the equations themselves; for a given point, it is straightforward to locate the epipolar line containing that point. The problem is to find a set of epipolar lines that will cover the whole image and introduces a minimum distortion.

4.4.2.2 *Epipolar Constraint Relationship*

In visual perception of depth from multiple views of a scene taken from different camera positions, the distance between the two images of the same real-world point can be used to calculate the disparity which is inversely related to the depth of the point. This is a correspondence problem. Using epipolar geometry, the search for a corresponding point can be reduced from searching over a whole image to just searching across a horizontal line of pixels in the image, called the epipolar line [310].

A world point P in two views p_1 and p_2 due to cameras C_1 and C_2 is considered as shown in Figure 4.1. The projection e_{12} of C_1 onto the image plane of camera with centre C_2 is an epipole which is expressed as

$$e_{12} = [B_2|b_2] \begin{bmatrix} C_1 \\ 1 \end{bmatrix} \tag{4.12}$$

where $\begin{bmatrix} C_1 \\ 1 \end{bmatrix}$ is the homogeneous coordinate of C_1. Using (4.4) in (4.12), we have

$$e_{12} = -B_2 B_1^{-1} b_1 + b_2 \tag{4.13}$$

By taking the depths μ_1 and μ_2, of P the from centre of the two cameras into consideration, the expressions in (4.2) and (4.3) can be written as

$$\mu_1 p_1 = [B_1|b_1]\, P = [B_1|b_1] \begin{bmatrix} P \\ 1 \end{bmatrix} \tag{4.14}$$

$$\mu_2 p_2 = [B_2|b_2]\, P = [B_2|b_2] \begin{bmatrix} P \\ 1 \end{bmatrix} \tag{4.15}$$

Equation (4.15) can be rewritten based on (4.14) as

$$\mu_2 p_2 = B_2 B_1^{-1} \mu_1 p_1 + \left(b_2 - B_2 B_1^{-1} \right) = B_2 B_1^{-1} \mu_1 p_1 + e_{12} \tag{4.16}$$

$$\mu_2 p_2 - B_2 B_1^{-1} \mu_1 p_1 - e_{12} = 0 \tag{4.17}$$

$$\begin{bmatrix} \mu_2 & B_2 B_1^{-1} \mu_1 & 1 \end{bmatrix} \begin{bmatrix} p_2 \\ p_1 \\ e_{12} \end{bmatrix} = 0 \tag{4.18}$$

Based on the linear dependency exhibited by (4.18), the equation can be written as

$$p_1^T \left(B_2 B_1^{-1} \right)^T e_{12} p_2 = 0 \tag{4.19}$$

$e_{12} = \left(e_x, e_y, e_z \right)^T$ can be expressed as a cross product matrix in the form

$$[e_{12}]_\times = \begin{bmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{bmatrix} \tag{4.20}$$

Therefore (4.19) can be written as

$$p_1^T F p_2 = 0 \tag{4.21}$$

where $F = \left(B_2 B_1^{-1} \right) [e_{12}]_\times$ is termed fundamental matrix and provides a representation of both the intrinsic and extrinsic parameters of the two cameras [311]. F is always of rank 2 [312]. This means one of its eigenvalues is always zero. Furthermore, the fundamental matrix F relates a point in one stereo image to the line of all points in the other stereo image that may correspond to that point according to the epipolar constraint. It is endured that the epipole in the second panoramic view is obtained as the right null space to the fundamental matrix and the epipole in the first image is obtained as the left null-space to the fundamental matrix. An important contribution of this work therefore is to get the stereo-panoramic images to satisfy this condition in order for an accurate depth map to be generated.

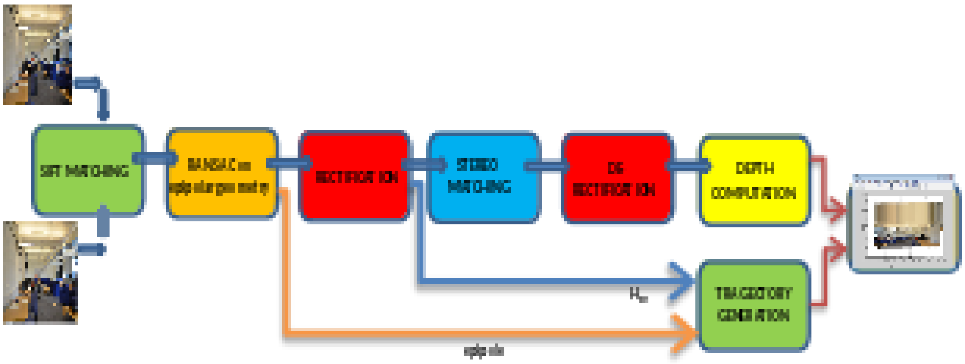


Figure 4.22: View synthesis pipeline.

It has to be observed that each point correspondence generates one linear equation in the entries of F. Given at least 8 point correspondences it is possible to solve linearly for the entries of F up to scale.

In this work, to generate a complete depth map, the solution of (4.21) is obtained by first normalizing the match points based on the knowledge of the intrinsic parameters of the acquisition camera. This helps to significantly reduce the effect of residual distortion. The camera intrinsic parameters are further optimised. Singular value decomposition [313] of F is then performed. This approach is relatively attractive since minimised frobenius norm can be easily obtained.

4.4.3 *View Synthesis Pipeline*

The view synthesis pipeline of [166] shown in Figure 4.22 has been adopted. Stereo-panoramic view is first generated from images samples acquired through camera panning. This is achieved through a series of steps namely, image acquisition, inverse cylindrical warping and correction of radial distortion, feature detection and matching, RANSAC motion estimation, and image blending and drift correction.

SIFT algorithm is used to extract features from each panoramic image. Each SIFT descriptor is 128-dimensional vector. The features from the two images are matched and RANSAC algorithm is ran to eliminating small number of outliers that has the potential to affect the accuracy of the generated fundamental matrix with a by-product called epipole.

The essence of rectification is to produce an image in which epipolar lines are parallel and horizontal and corresponding points have the same vertical coordinates [23]. Infinite plane homography $H_{\infty 12}$ is also obtained as a by-product of rectification.

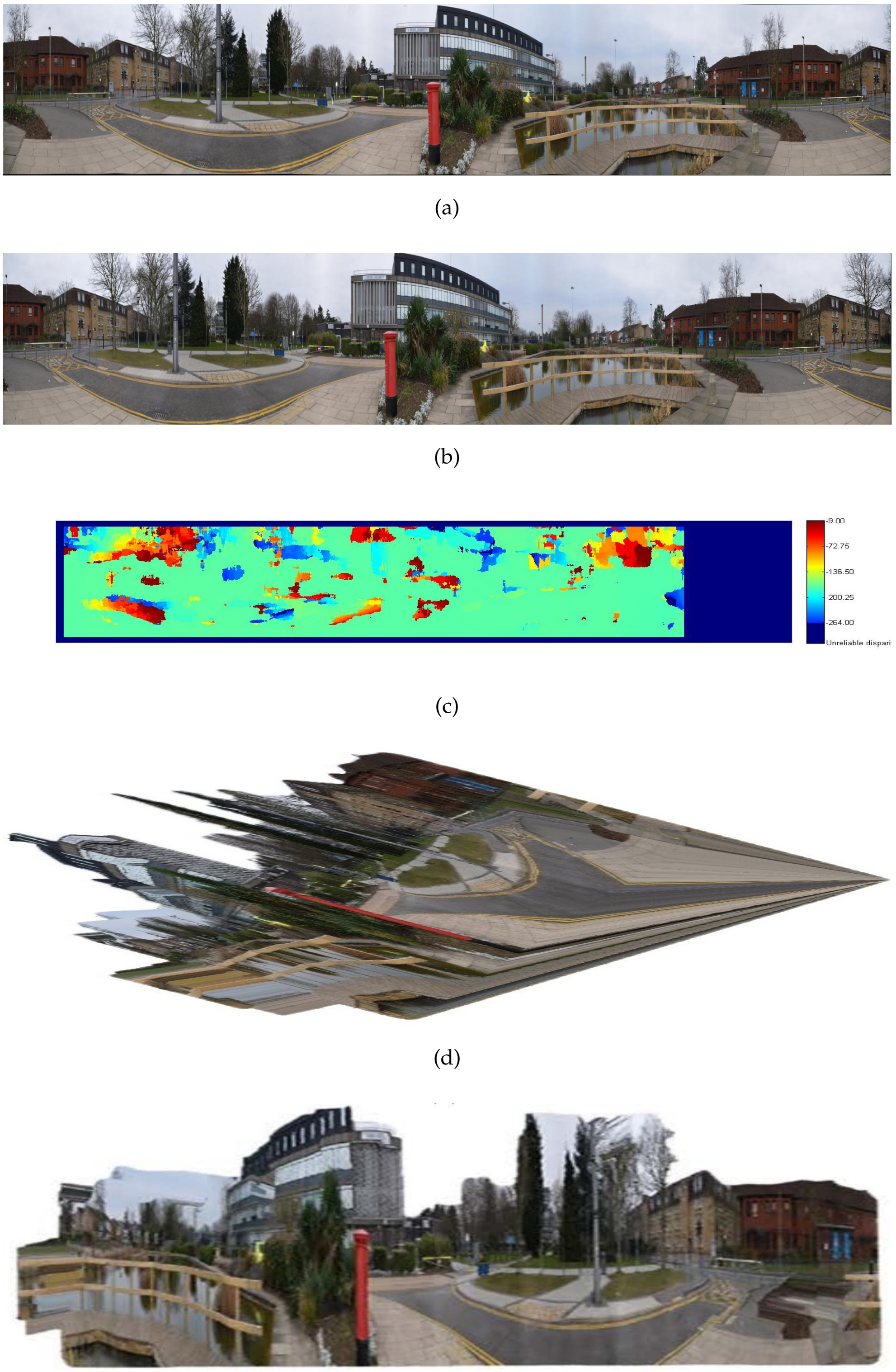
Dense stereo matching is performed using the rectified images. De-rectification process is carried out which allows for the computation of the Z dimension of the scene. Knowing the coordinate of a virtual viewpoint, a virtual view of the original panoramic image can be calculated.

4.4.4 *Results and Discussions*

In Figure 4.23(a) and (b) is a stereo-panoramic view of the scene used in this experiment. Each panoramic view contains thirty eight image samples acquired with a compact camera. The depth map common to the stereo-panoramic view is in Figure 4.23(c). The virtual environments generated by texturing the depth map are shown in Figure 4.23(d) and (e) respectively. Because the stereo-panoramic view represents an outdoor scene with a large depth of view, the shapes of the structures are not so conspicuous in the depth map even though the correct information is contained therein. It can be clearly seen that the two generated virtual views are considerably different from the initial stereo-panoramic views.

4.4.5 *Objective Analysis*

The objective assessment of the results in this section is based on **QoE!** (QoE!) discussion presented in the appendix. Detail information about the quality assessment of the rendered images can be found in Figure 4.24 and Figure 4.25. Each figure provides information about the edge profile and MTF parameters. The edge rise distance (10-90%) expressed in pixels is 188.75 and 126.35 pixels respectively. This parameter has a larger value in Figure 4.24 compared to its value in Figure 4.25. These figures also feature an undershoot of 2.1% and 4.7% respectively.



(a)

(b)

(c)

(d)

(e)

Figure 4.23: Two panoramic images which represent left and right image samples are shown in (a) and (b). The constructed depth map is shown in (c). Two synthesised virtual views are shown in (d) and (e).

However, in Figure 4.24 and Figure 4.25, sharpness which is the most important image quality factor is measured in terms of contrast represented by MTF. The frequency where contrast falls to 50% of its low frequency value, which corresponds well with perceived image sharpness is about twice the Nyquist frequency in Figure 4.24 and twenty eight times in Figure 4.25. Nyquist frequency is a measure of the degree of aliasing. In Figure 4.24, the region of of interest which is automatically determined by the imatest software is around the texture areas and has a size of $521 \times 425\, piexels$. MTF is low at low spatial frequencies. It is only high when the spatial frequency is fast approaching $50 cycles/mm$. Of course, in this case, MTF is low for low spatial frequency, the images still has good quality subjectively.

4.4.6 *Conclusion*

The photographic computation of virtual environment in software from stereo-panoramic view has been presented. The implementation involves three basic steps namely: generation of stereo-panoramic view, construction of depth map, and depth map texturing. The implemented algorithm ensures that the centre of one camera is located in the image of the other, hence satisfying epipolar constraint which also yields a fundamental matrix of rank two and whose determinant is zero. These subtle parameters indicate the possibility of depth map generation. Another strength of this work is that the synthesized virtual environment contains structures from the initial stereo-panoramic view that have been either repositioned or reshaped.

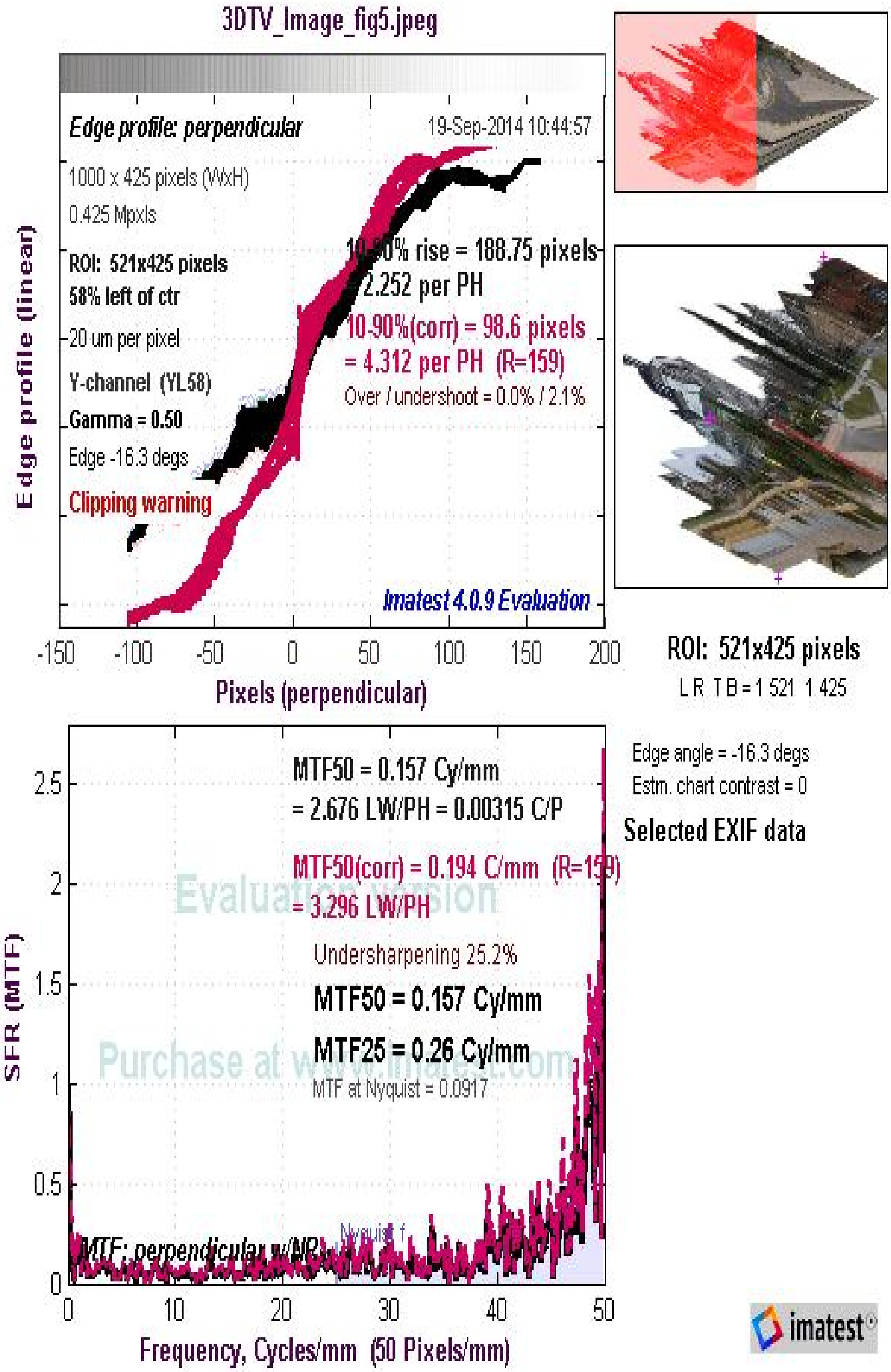


Figure 4.24: Image objective evaluation using Edge profile and MTF performance.

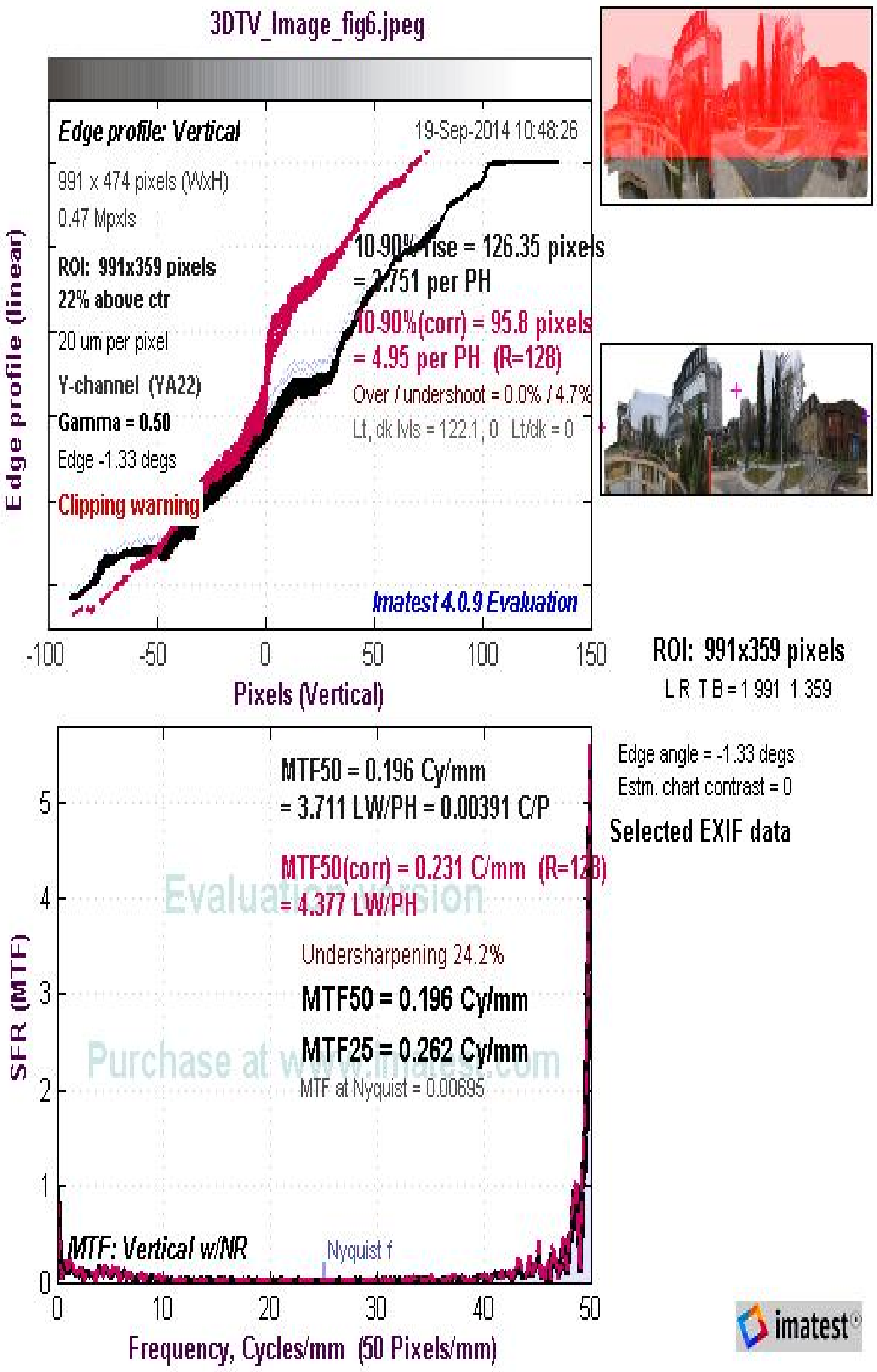


Figure 4.25: Image objective evaluation using Edge profile and MTF performance.

5 METRIC RECONSTRUCTION

5.1 INTRODUCTION

The central theme of this chapter is the reconstruction of a scene from multiple images with SCOP. In the literature related to computer vision, scene reconstruction from multiple images with multiple centre of projection is common. SCOP is an important feature of image acquisition for use in the construction of panoramic images. A more fundamental justification is that the use of images acquired based on SCOP in 3D reconstruction poses serious challenges with regards to the accuracy of the reconstructed 3D points. Hence, the need to optimise such points becomes crucial. The important consequence is the use of SBA algorithm which optimises the reconstructed 3D points remarkably well. First reconstruction of synthetic scenes is studied. Then the reconstruction of real scenes is considered with and without the use of SBA. The implementation is performed in Matlab while scene visualisation is done with Meshlab software.

5.2 RECONSTRUCTION OF SYNTHETIC SCENE

Here, synthetic scene implies using a set of randomly generated points to represent a scene. This idea emanates from the fact that any worldly object or structure is comprised of a set of points. 3D reconstruction can be defined as the problem of using 2D measurements arising from a set of randomly generated points depicting the same scene from different viewpoints, aiming to derive information related to the 3D scene geometry as well as the relative motion and the optical characteristics of the camera(s) employed to acquire these images.

To demonstrate the idea of scene reconstruction, three fundamental steps are outlined. 1) These are creation of scene points and camera pose, 2) superposition of error on the scene and pose data, 3) and data recovery through bundle adjustment technique.

The implementation of the first step starts with the assumption of a certain number of cameras of known intrinsic parameters and image resolution. The poses of the assumed cameras are generated in Matlab based on normally distributed pseudorandom numbers. Mersenne twister (default), Multiplicative congruential generator, Multiplicative lagged Fibonacci generator, Multiplicative lagged Fibonacci generator, Shift-register generator summed with linear congruential generator, and Modified subtract with borrow generator are the random number generators available in Matlab. Mersenne twister has been used in this work.

Mersenne Twister provides a super astronomical period of $2^{19937} - 1$ and 623-dimensional equidistribution up to 32-bit accuracy, while using a working area of only 624 words [314]. This is a new variant of the previously proposed generators, Twisted Generalised Feedback Shift Register (TGFSR), modified so as to admit a Mersenne-prime period. Two new ideas have been added to the existing twisted Generalised Feedback Shift Register (GFSR). One is the realisation of Mersenne-prime period using incomplete array. Second, the primitivity of the characteristic polynomial of a linear recurrence can be tested with a fast algorithm. This is called the inversive-decimation method. The characteristic polynomial has many terms.

450 2D Normally Distributed Pseudorandom Pixels Locations (NDPPL) with corresponding depths are generated on the image plane of the reference camera. 450 is expected maximum number of features to be tracked on each frame of 640×480 resolution. Using the chosen reference camera (first camera), the pixel location are re-projected so as to determine the 3D point. Only pixel locations with positive depth are re-projected and registered in the $Number\,of\,scene\,points \times number\,of\,cameras$, visibility map. The generated 3D points are tested to see if they can be seen in the second camera. Since the number of pixels seen by the

second camera is less than the maximum expected, another set of NDPPL is generated based on the difference between the maximum expected to be tracked and number already tracked. The new set of NDPPL is again tested for positivity of depth and the pixels with positive depth are re-projected in order to generate their corresponding 3D points. This process is repeated until all the cameras have been considered. The total number of 3D points generated through re-projection of pixel locations in the process just described constitute the scene points. This is a reliable method which is widely used and also documented in literature.

Next, the scene points, camera pose, camera matrix, and pixel locations are corrupted with Normally Distributed Pseudorandom Error (NDPE). A statistical error is usually given by the standard deviation ρ, equal to the square root of variance (expectation value of the square of the difference to the mean).

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\, dx \tag{5.1}$$

where x is a measure of data. μ is the mean of the data set.

In the third step, an attempt is then made to recover scene points and camera parameters using the bundle adjustment. In the ordinary bundle adjustment, advantage is not taken of the existence sparsity. The implementation of ordinary bundle adjustment involves error minimisation by way of optimisation of parameters purely performed based on Matlab coding.

A simulated experiment based on the mentioned steps was conducted using Matlab. A set of scene points was created as shown in Figure 5.1 (b). The points are represented with a circle in red, blue, and yellow colours with respect to some cameras symbolically represented with triangles. Fifty cameras of known calibration were used to capture these scene points from different observation points. Each frame has a resolution of 640 × 480. In order to demonstrate the concept of scene reconstruction, the coordinates of the scene point, camera calibration matrix, rotation and translation were corrupted with some randomly generated error. Bundle adjustment was then used to reconstruct the scene points and camera parameters (intrinsic and extrinsic). Figure 5.1 (a) provides information regarding the number of scene points that can be seen in each camera frame.

The motion of the cameras from one point to another usually comprises of rotation and translation. Therefore, Figure 5.2 and Figure 5.3 present the recovered rotation and translation of the cameras. Rotation is about the x, y and z axis. Translation of the camera is in the x, y, and z directions. The recovered intrinsic camera parameters are presented in Figure 5.4. The intrinsic parameters appear as straight lines since they have the same values for all the cameras (cameras of the same model).

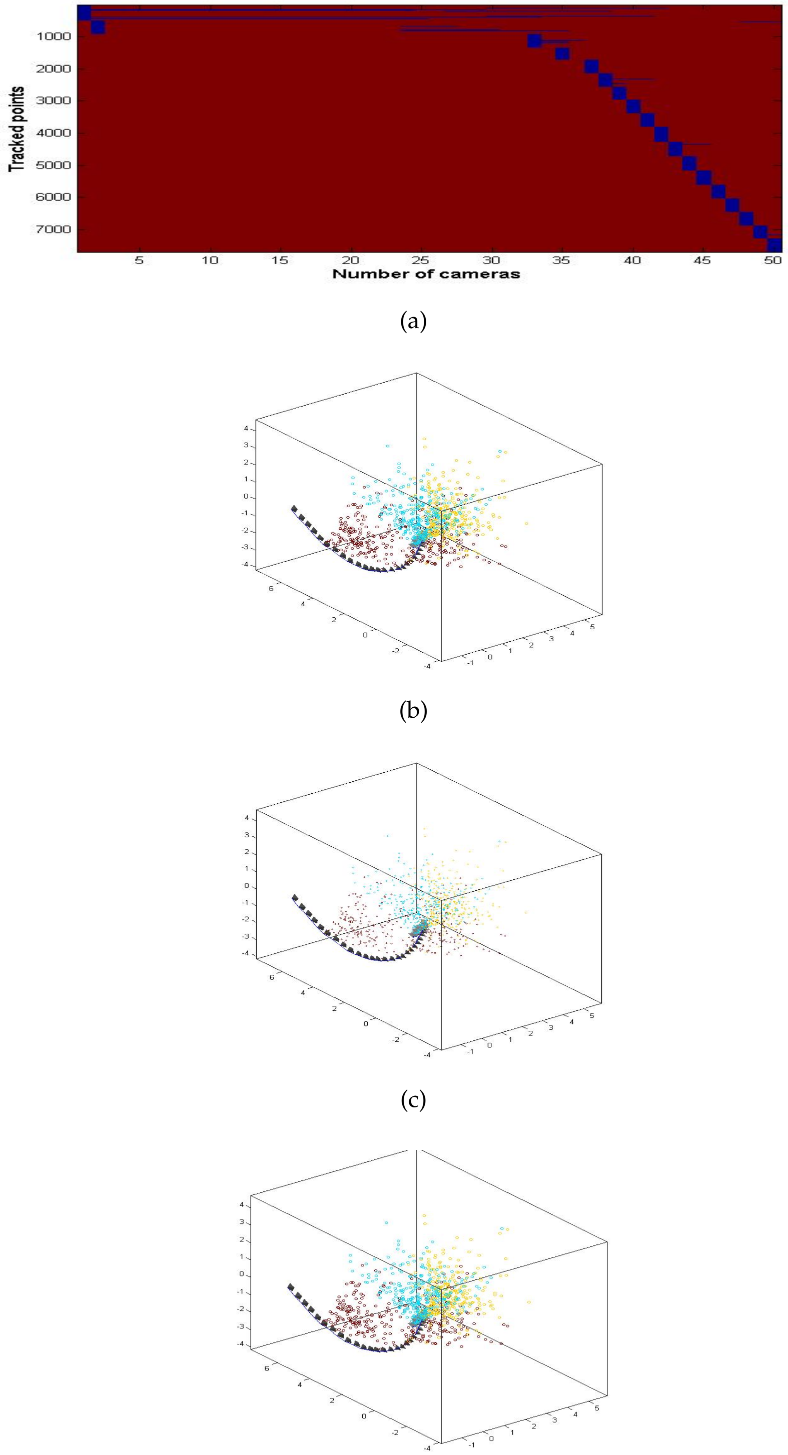
Figure 5.5 is another set of scene points that was created in simulation to explain the concept of bundle adjustment. The rotation error evaluation simulation is shown in Figure 5.6.

The same experiment is repeated using four 2D images. Figure 5.7 (a) shows two of the four 2D images with the extracted feature points superimposed on then. In Figure 5.7 (b), the re-projection of some of the feature points is shown. The re-projected feature points are shown with respect to the acquisition cameras in Figure 5.7 (c). Feature point depth, camera rotations and translations, and re-projection error are depicted in Figure 5.7 (d). The re-projection error is considerable because of the presence of outliers.

5.3 RECONSTRUCTION OF REAL SCENE USING BA

In this section the reconstruction of a scene is attempted from multiple images containing different optical and geometric properties of the same scene. Also, all image samples shown in Figure 5.8 have the same centre of projection. The steps that have been implemented in the reconstruction process are discussed as follows.

Feature detection, extraction and matching using SIFT are performed. This leads to the determination of fundamental matrix F that relates the two views. This is expressed in (5.2). It is singular with a rank of two and has seven degrees of freedom.



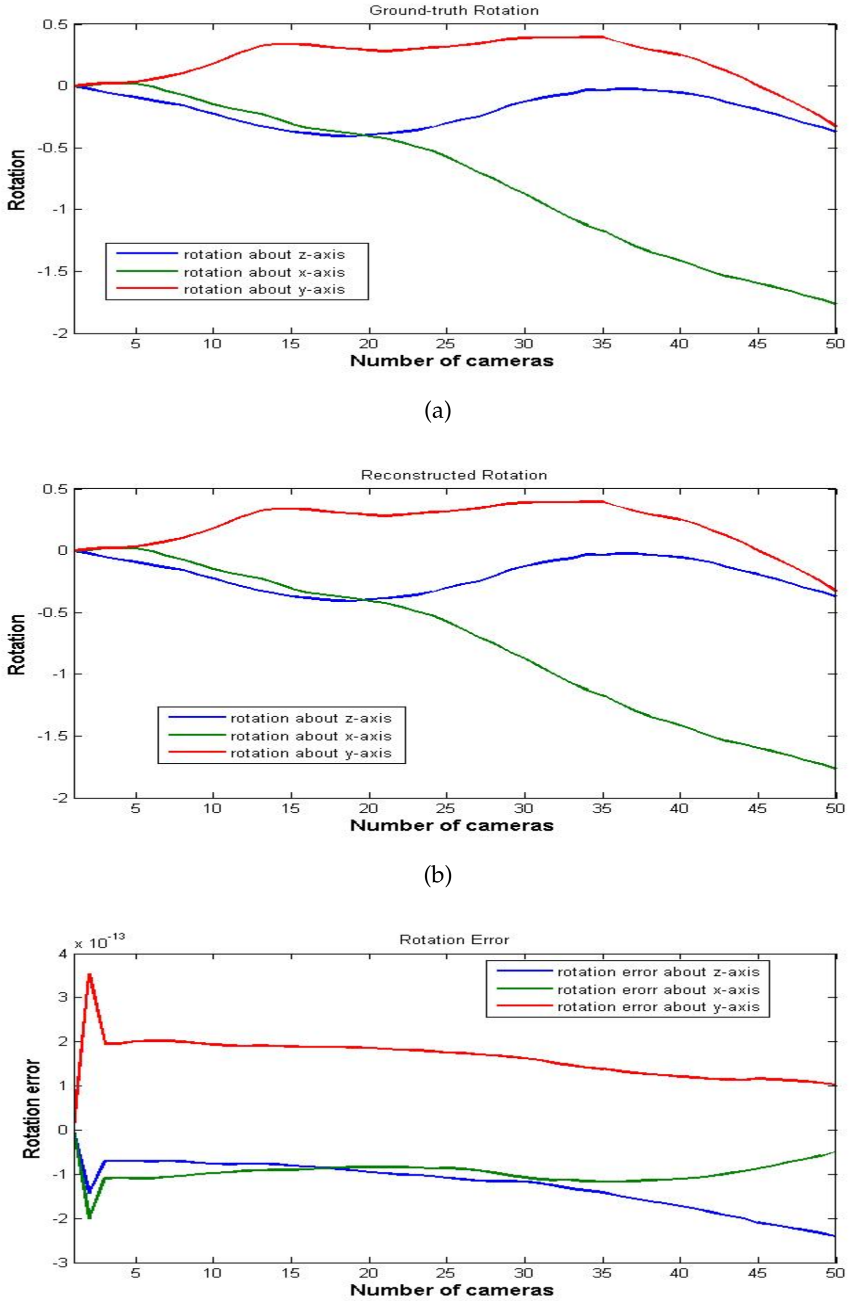
(a)

(b)

(c)

(d)

Figure 5.1: Demonstration of bundle adjustment. (a) shows the number of scene points captured by each camera. (b) shows a set of points representing a synthetic scene. (c) presents the corrupted scene. (d) shows the reconstructed scene.

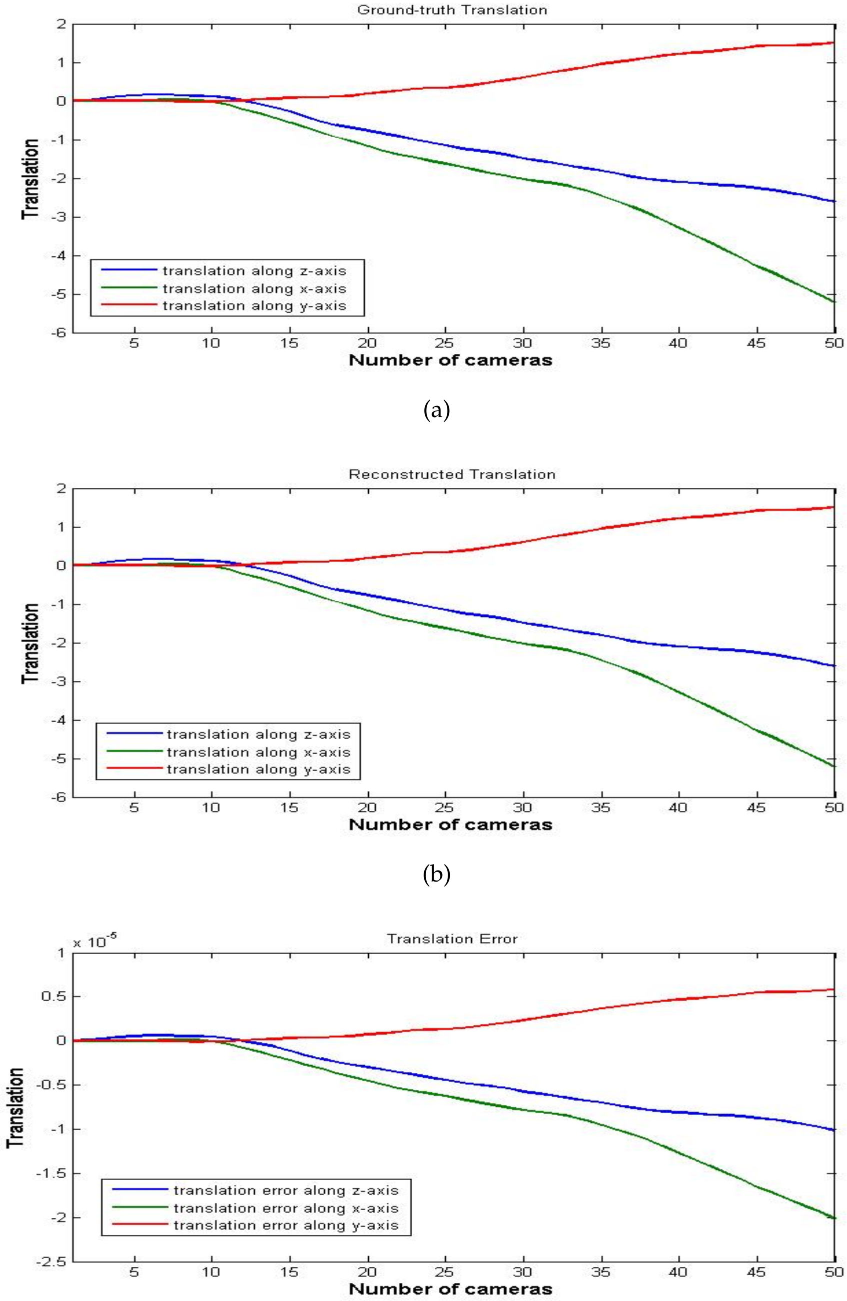


(a)

(b)

(c)

Figure 5.2: Demonstration of bundle adjustment. In this figure is shown the camera rotation about x, y, and z axis: (a) ground-truth rotation, (b) camera rotation required to cope with the corrupted of scene, (c) computed camera rotation error.

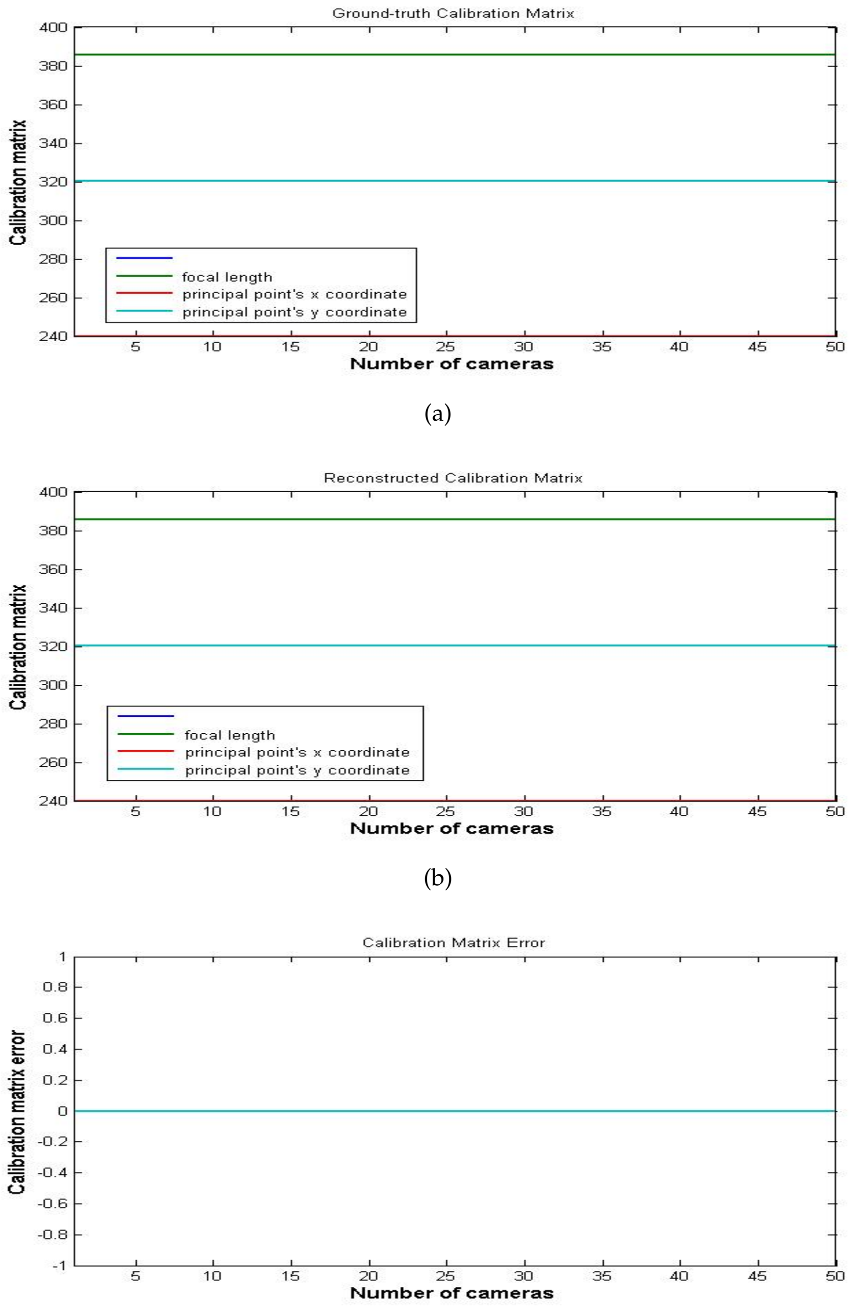


(a)

(b)

(c)

Figure 5.3: Demonstration of bundle adjustment. camera translation is presented: (a) camera ground-truth translation, (b) translation as a result of corruption of scene, and (c) translation error.

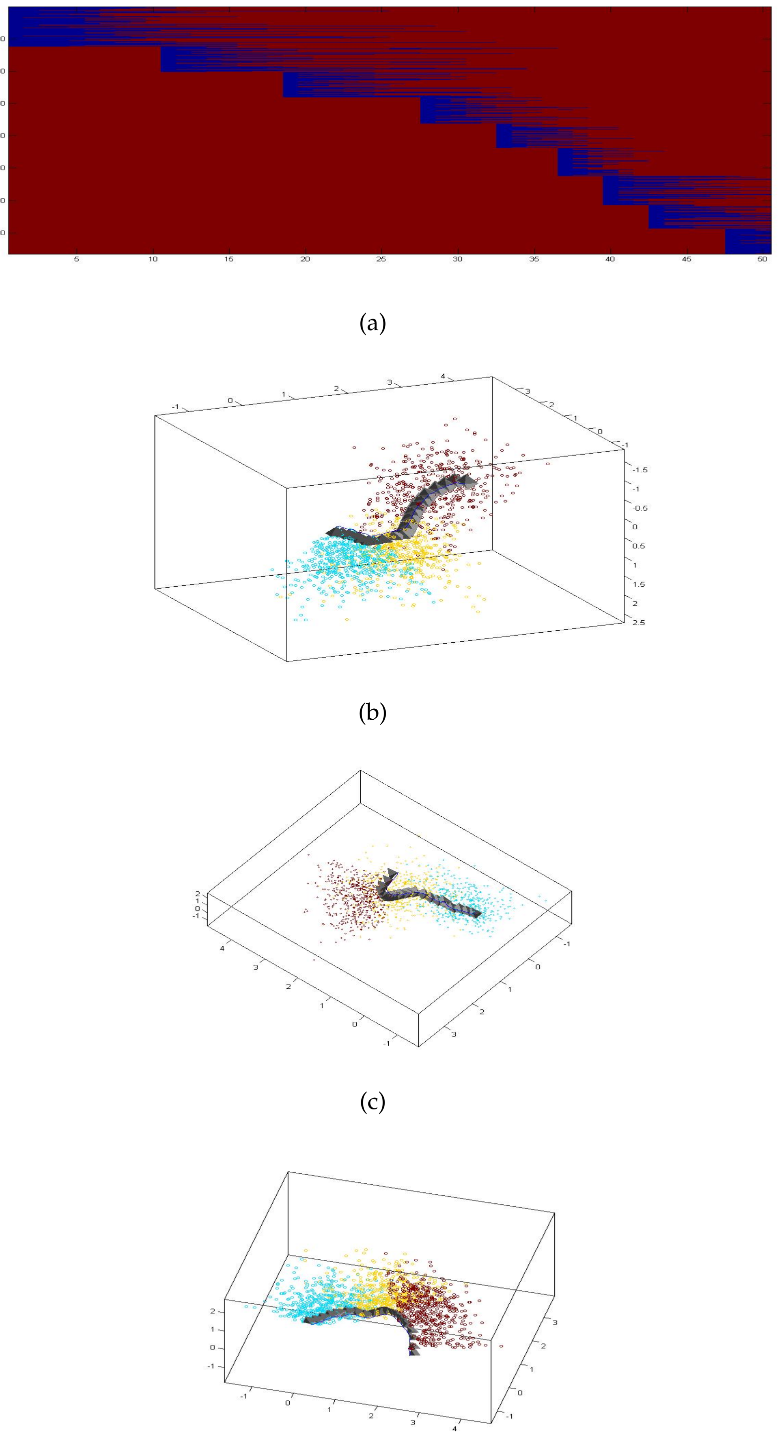


(a)

(b)

(c)

Figure 5.4: Demonstration of bundle adjustment. camera calibration matrix: (a) ground-truth calibration matrix, (b) calibration matrix as a result of corruption of scene, (c) calibration matrix error. A straight line plot means the intrinsic parameters are the same for all the cameras used.



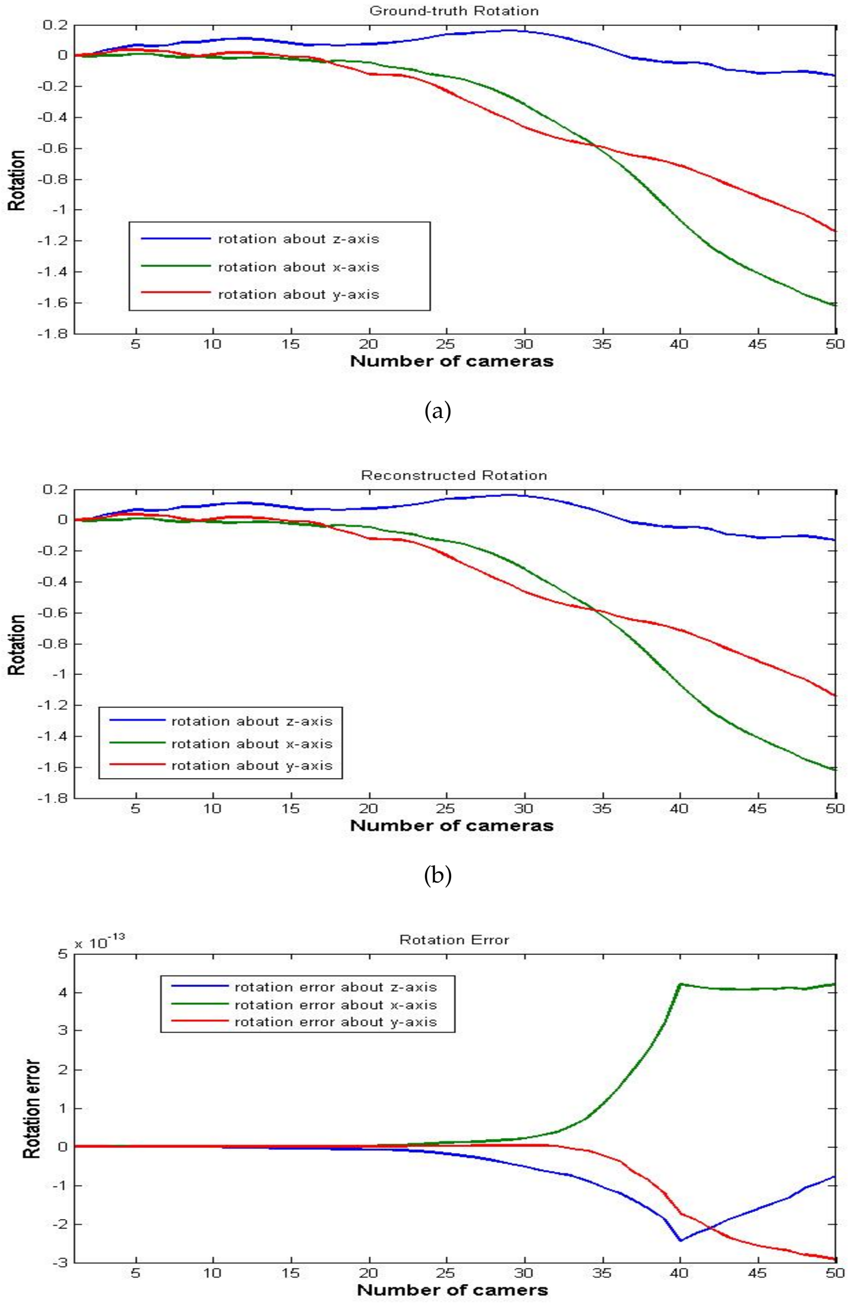
(a)

(b)

(c)

(d)

Figure 5.5: Reconstruction of synthetic scene. (a) shows the number ofscene tracked in each camera frame, (b) is a synthetic scene, (c) corrupted scene, (d) reconstructed scene.

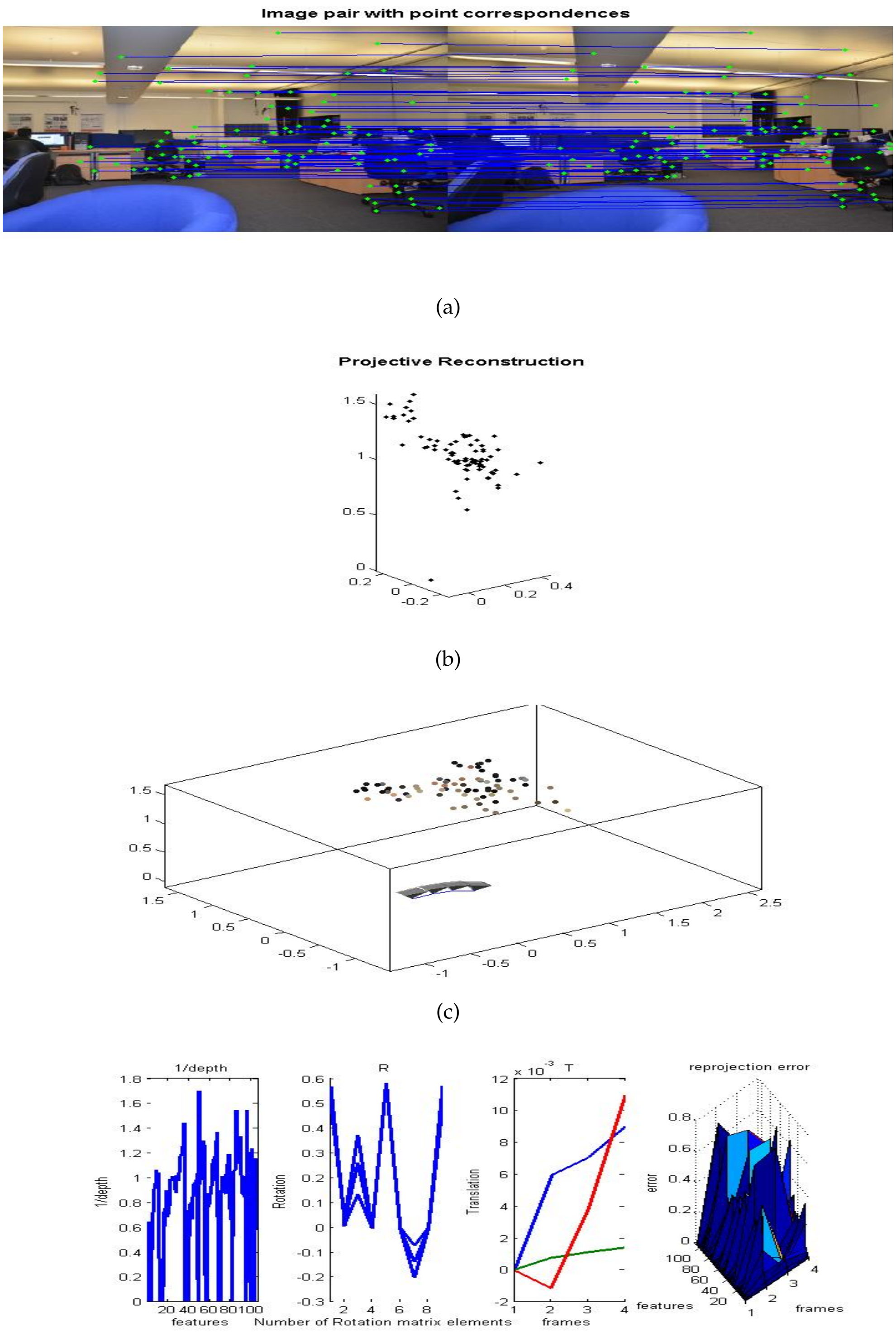


(a)

(b)

(c)

Figure 5.6: Rotation error evaluation for synthetic scene shown in Figure 5.5: (a) ground-truth rotation, (b) rotation due to corruption of scene, and (c) rotation error.



(a)

(b)

(c)

(d)

Figure 5.7: Reconstruction of feature points from 2D images. (a) shows image pair with corresponding points. In (b) is the projective reconstruction. (c) shows the scene with respect to camera. Finally, (d) shows performance evaluation comprising of depth, rotation, translation, and re-projection error.

$$F \simeq K^{-1} S\left(.\right) RK. \tag{5.2}$$

where $S\left(.\right)$ is the skew-symmetric matrix which defines translation in terms of three components, R is the rotation matrix, and K is the camera matrix made of intrinsic parameters.

Using the computed fundamental matrix, the essential matrix E, is estimated. E is then decomposed to allow triangulation and pose determination process to be carried out.

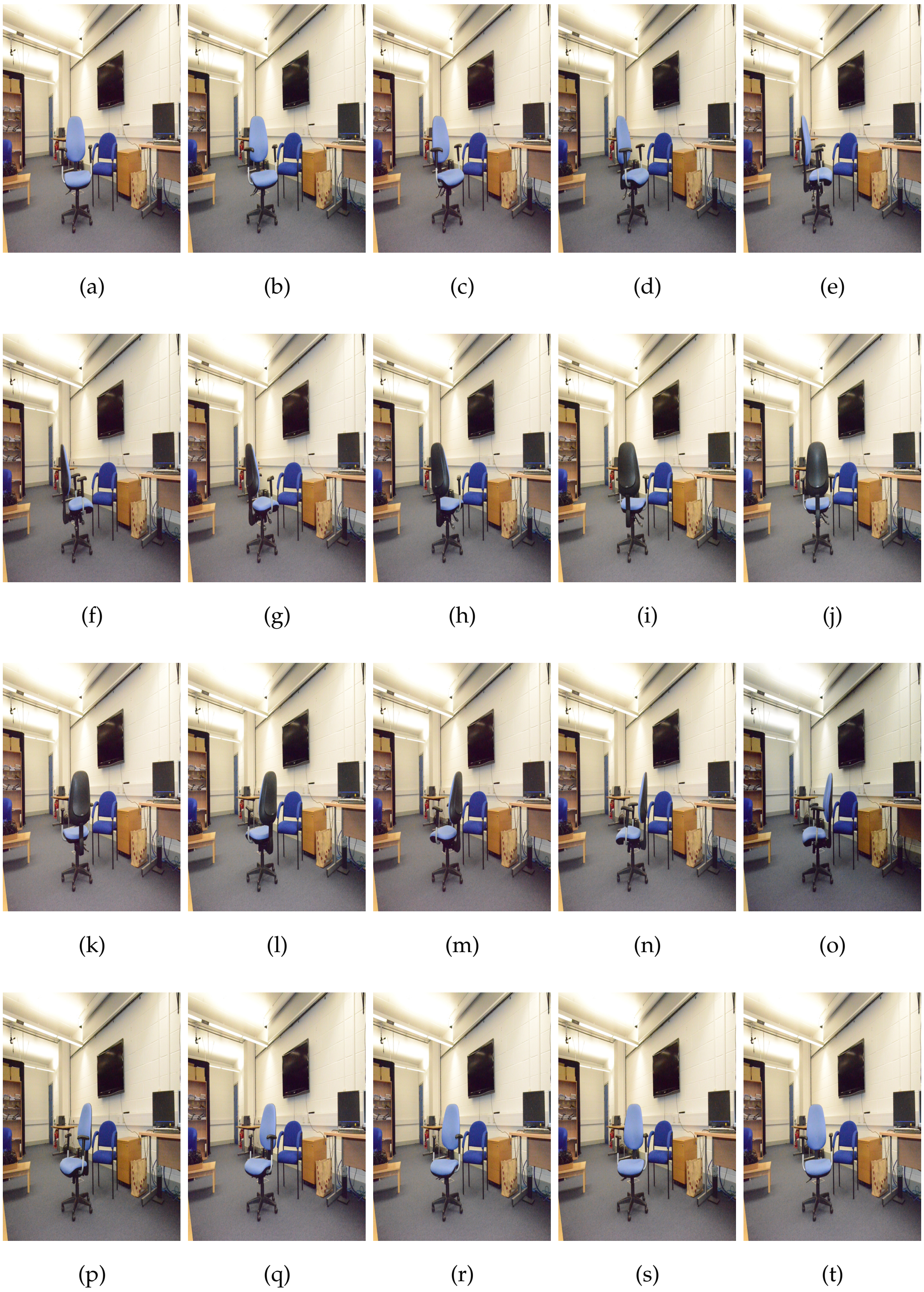
$$E \simeq K^{T} FK. \tag{5.3}$$

Track computation is the third step. From the result of the triangulation process, the 3D points visible in each image are generated as tracks.

Optimisation of tracked 3D points and camera parameter is done using bundle adjustment. The process actually involves running an executable file ba2D [315].

Using the grey images in Figure 5.9, dense stereo for all the image samples is now calculated. Starting with the optimised camera poses for any two of the images, 3×4 projective matrices are calculated. The earlier tracked 3D points are then reacquired based on a new computed and optimised projective matrices. The difference between the current pixel coordinates and the earlier ones before bundle adjustment is applied constitutes the re-projection error [315].

Generation of a ply file containing the reconstructed vertices and faces is the final step. PLY is a computer file format known as the Polygon File Format or the Stanford Triangle Format [316]. The need to store three-dimensional data from 3D scanners is the main purpose behind the idea of this file format type. In this file format, an object is represented as a list of nominally flat polygons. Colour and transparency, surface normals, texture coordinates and data confidence values are the known properties commonly stored in ply files. It is also said that different properties for the front and back of a polygon can be handled in this



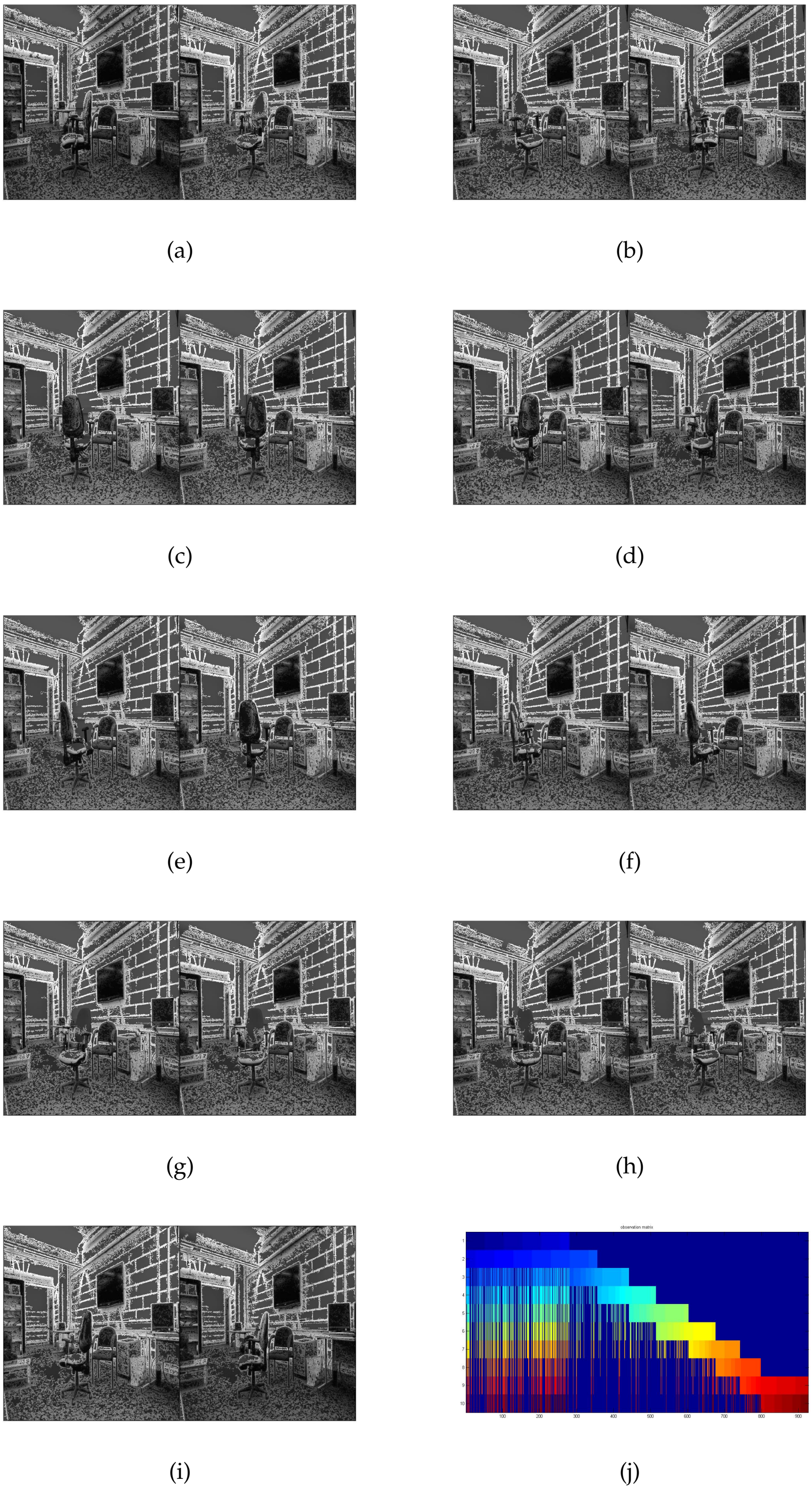
(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

(p) (q) (r) (s) (t)

Figure 5.8: Image samples with a single centre of projection used in this work.



(a) (b)

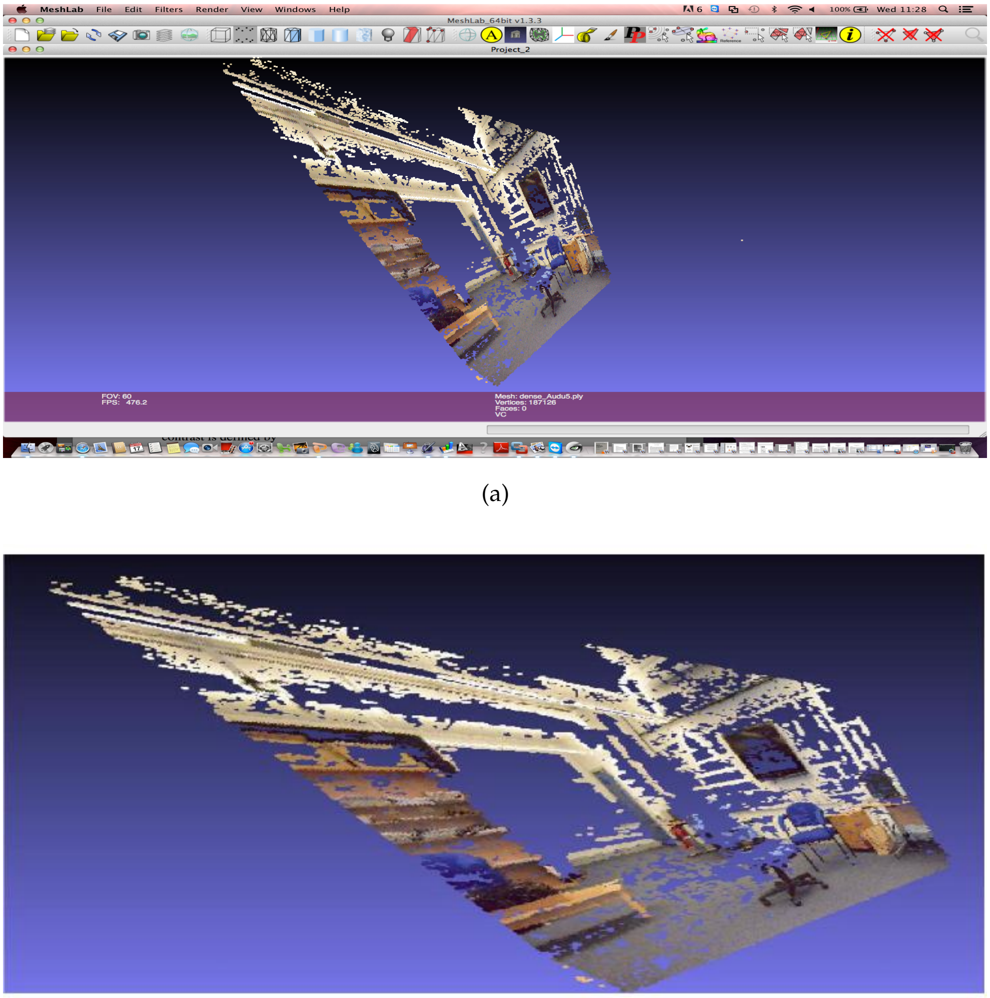
(c) (d)

(e) (f)

(g) (h)

(i) (j)

Figure 5.9: Gray images used in feature extraction are shown in (a) through to (i). The number of 3D points seen in each frame is shown in (j).



FOV: 60
FPS: 476.2

Mesh: dense_Audu5.ply
Vertices: 187126
Faces: 0
VC

(a)

(b)

Figure 5.10: Reconstructed scenes from multiple images with the same centre of projection. (a) actual reconstructed scene, (b) zoom-out form of (a).

format. The visualisation of the 3D points contained in the ply file is shown in Figure 5.10.

5.4 RECONSTRUCTION OF REAL SCENE USING SBA

The task of scene reconstruction from the multiple images shown in Figure 5.8 is the main consideration of this section. It is carried out through the use of SBA a C/C++ implementation of generic bundle adjustment based on the sparse Levenberg-Marquardt algorithm [317]. SBA is implemented using Linear Algebra PACKage (LAPACK) written in Fortran 90. LAPACK routines are written so that as much as possible of the computation is performed by calls to the Basic Linear Algebra Subprograms (BLAS). It provides routines for solving least-squares solutions of linear systems of equations [219].

In order to use SBA, the images have to be understood contextually. Context is said to be a consequence of theories about cognitive processes, not something observed. It has a theoretical role, not one of a measurable unit. The use of contextual constraints is considered to be indispensable for a capable vision system. It is observed that contextual constraints are ultimately necessary in the interpretation of visual information [318]. It therefore means the scene contained in the image samples is best understood in the spatial and visual context of the objects in it; the objects are recognised in the context of object features at a lower level representation; the object features are identified based on the context of primitives at an even lower level; and the primitives are extracted in the context of image pixels at the lowest level of abstraction [319].

Markov random field MRF theory provides a convenient and consistent way for modelling context dependent entities such as image pixels and correlated features. This is achieved through characterising mutual influences among such entities using conditional MRF distributions. The practical use of MRF models is extensively demonstrated in [320] in which a scene is segmented into regions based on texture.

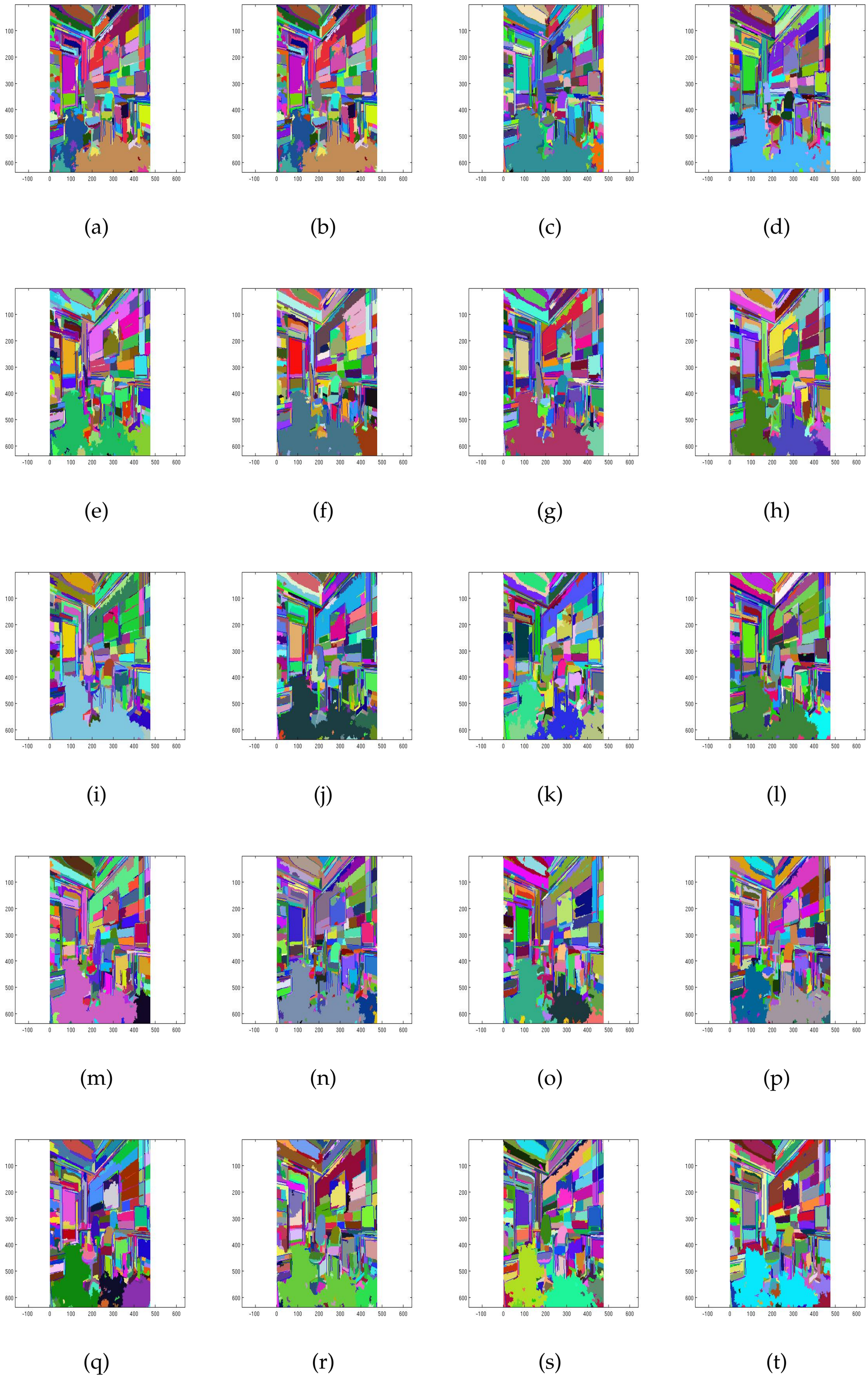
Hence, in this work, segmentation of sample images and other steps are implemented in Matlab based on the idea discussed in [321–324]. The segmentation result for the different image samples are shown in Figure 5.11. Consequently, feature detection, extraction, and matching is performed on segment bases rather than the whole image at a goal.

The other implementation steps are similar to that mentioned in section 5.3 except that the optimisation process is performed using generic SBA [219]. This means that starting with initial estimates of the 3D coordinates corresponding to a set of points seen in image samples, as well as initial estimates for the viewing parameters pertaining to each image as explained in section 5.3, SBA is a large optimisation problem which take advantage of the sparsity associated with the linear equations to be solved. It involves the simultaneous refinement of the 3D structure and viewing parameters (i.e. camera pose and possibly intrinsic calibration and radial distortion), aiming to obtain a reconstruction which is optimal under the assumption of zero-mean Gaussian regarding the noise pertaining to the observed image features. So long as outliers have been removed from the initial estimate of 3D points using geometry-based techniques, SBA algorithm can demonstrate a high degree of robustness.

At the heart of SBA algorithm is eucsbademo.c. This visual C file is modified in two ways. First it is modified so that motion and/or structure parameters are printed to an output file after the SBA has been ran. Second an assignment to perform motion and structure optimisation is set. The compilation of eucsbademo.c and other related files are then performed. This step allows some of the necessary libraries to be created.

Whenever eucsbademo.c is ran, metric reconstruction [325] is performed using the optimised 3D points and camera parameters. During this process, mono estimation, robust pose and match generation, dense match, re-inference, coherence texture processing, volumetric refining, and rendering steps are implemented [322].

The numerical results of SBA simulation is output to a file. This is as a result of the modification mentioned earlier on. The file not only contain statistics regarding the minimisation but also 3D coordinates of scene points, camera pose, and image pixel coordinates are. The statistics regarding the minimisation on one



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

(q) (r) (s) (t)

Figure 5.11: Segmented image samples.

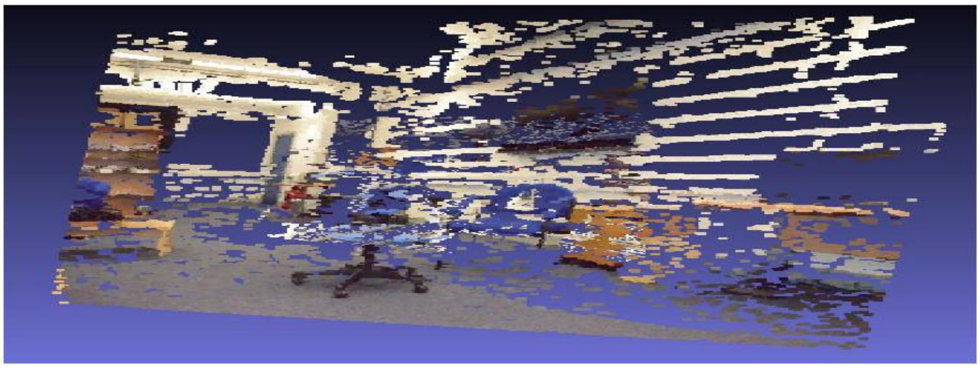


Figure 5.12: Scene reconstruction based on SBA.

of the image pairs is shown in sections A.3. In this section, the four first four lines summarises the statistics regarding the minimisation that has been performed. It describes the number of 3D points that have been optimised, number of cameras or frames considered, number of image pixels, SBA method used i.e structure and motion (BA_MOTSTRUCT), expert driver, analytic Jacobian, fixed intrinsics, and whether or not covariance is used. It also gives a summary of the number of iterations before convergence is achieved. The initial and final errors are also described. In section A.3, for example, SBA algorithm reduces the error from 1235.07 to 0.0108958 in 200*msec*.

Usually, in the output file, the motion parameters which describe the poses of the cameras are given in line six and seven. The pose of the reference camera is given in line six and consists of four elements quaternion and three elements translational matrix. Line seven contains the pose of the second camera.

The rendered 3D scene is visualised in Meshlab environment and is shown in Figure 5.12. The quality of the rendered scene is not very good as should be expected from SBA. There is the presence of holes. This is purely due to two fundamental reasons. First image samples with SCOP have been considered in this work since it is applicable to panoramic images. The second reason is related to the inaccuracy embedded in the initial estimate of 3D parameters. However, when the reconstructed image is zoomed into, the size of the observed holes is reduced.

5.5 CHAPTER SUMMARY

Metric reconstruction from multiple 2D representation of the same scene have been investigated. The investigation started with a corrupted randomly generated scene points in order to highlight some of the critical challenges involved in scene reconstruction. This study then progressed to the use of multiple image samples in real scene reconstruction using bundle adjustment without necessarily taking advantage of sparsity. The obtained results are reasonable.

The third experiment involves the use of SBA which in theory is expected to perform computationally better because of the sparsity associated with it. The obtained numerical results actually pointed to this based on the statistics regarding the minimisation that is performed through the use of SBA. The main challenge encountered is the visualisation of the optimised numerical results in Meshlab software environment. The optimised 3D points have been rendered with some holes present. Understandably, this is attributable to the fact image samples with SCOP have been used and the presence of minute number of outliers in the initial estimate of the 3D points parameters.

6 TRAPEZOIDAL ARCHITECTURE

6.1 INTRODUCTION

Visual content acquisition is a strategic functional block of any visual system. Despite its wide possibilities, the arrangement of cameras for the acquisition of good quality visual content for use in multi-view video remains a huge challenge and a further area of development. This chapter examines a new camera topology which can be used for multi-view video. The details are presented which gives the description of trapezoidal camera architecture and relationships which facilitate the determination of camera position for visual content acquisition in multi-view video, and depth map generation. The strong point of Trapezoidal Camera Architecture is that it allows for adaptive camera positioning by which points within the scene, especially the occluded ones can be optically and geometrically viewed from several different viewpoints either on the edge of the trapezoid or inside it. Trapezoid characteristics, and the fact that the positions of cameras (with the exception of few) differ in their vertical coordinate description could very well be used to address the issue of occlusion which continues to be a major problem in computer vision with regards to the generation of depth maps. The contribution of the chapter is that a trapezoid graph is used in the formulation of the concept of trapezoidal camera architecture. Different geometries which can be used to determine camera viewpoint are discussed. This idea is then used to construct viewpoint dependent depth map of the scene object in Matlab and Maya environments. Significantly, depth map is one of the requirements for view synthesis in DIBR. The construction of depth map is based on part of the IBR implementation strategy presented in chapter four. In Maya, the

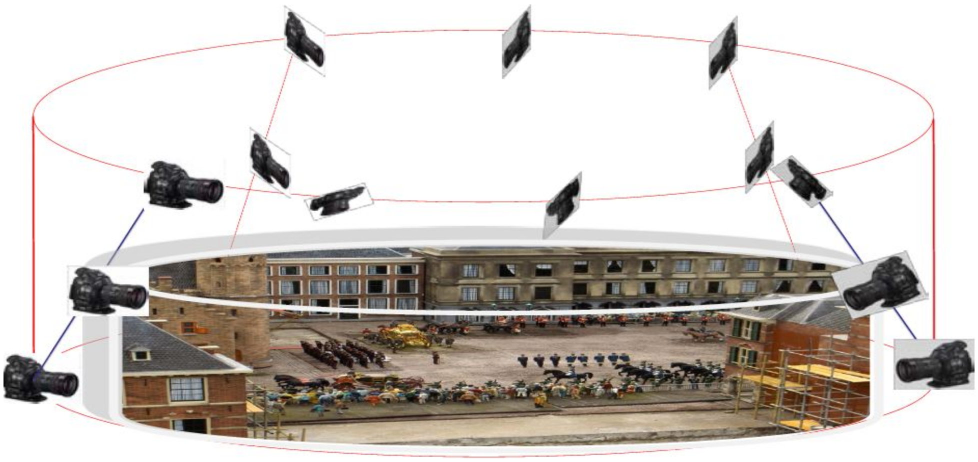


Figure 6.1: Pictorial representation of scene and trapezoidal camera architecture.

ability to view the scene from different viewpoints on the edge of a trapezoid is demonstrated.

6.2 BACKGROUND

A substantial body of evidence according to [326] demonstrates that exposure experience allows the individual to learn a great deal about the stimulus object, so that the ability to recognise, discriminate, and categorise the object generally improves. The first order step in the achievement of these improvements is the registration of both optical and geometric properties of the scene. The identification of this knowledge has triggered the visual content acquisition of scenes from a single viewpoint using optical cameras.

TCA demonstrated in Figure 6.1, could facilitate a change in the emphasis for the realisation of virtual camera. Of course, this computational photography realised in software provides a unique way to reduce the number of physical cameras used in the acquisition of visual content. A critical concern is the determination of the stereoscopic distance between a virtual viewpoint and any physical camera around it.

Also the analysis of TCA boils down to a search problem which entails the determination point coordinates. Since all the cameras are not at the same horizontal level, the problem of occlusion has a great chance of being addressed. It also provides an exponential population of possible viewpoints based on the properties of trapezoids as in [327], from which any points within the scene can be seen. This approach has the tendency to dramatically address the issue of occluded parts of an object even though the technique of synthetic aperture focusing is a design space that has been used in [328]. Therefore, the potential of TCA is explored in the generation of high quality depth map with respect to different virtual viewpoints.

It is worth noting that the trigonometric, area, sides and distances, and collinearity characteristics of this architecture very easily allow the baseline separation between any two cameras to be accurately computed. This parameter is important since it is the sole determinant in the categorisation of camera topology as either dense or sparse.

TCA is worth considering since a strong argument [328] has been put forward in favour of radially captured images. It is stated that linear array of cameras is considered to be the more correct method of capturing stereo images and does not suffer from distortion. However, convergence camera arrangement is simpler to implement: standard cameras or renderers can be used with no modification. Also, with respect to depth map acquisition, optimisation of matching energy function defined by MRF can now be optimised. This efficiently provides for rectification of the image pairs acquired through convergence camera array [329].

One eminent merit which triggers a very strong line of discussion in support of TCA is driven by the concept on which concentric mosaicking is based, as depicted in Figure 6.2. If a camera is rotated anti-clockwise along the circumference of a circle, a corresponding sinusoidal path, whose frequency depends on the speed of the camera, is generated. It therefore means that the reverse process whereby a camera translates along a sinusoidal path should be capable of realising a concentric mosaic. Since an efficient trapezoid is a truncated or clipped sinusoid, it does intellectually make sense in the face of conceptual and

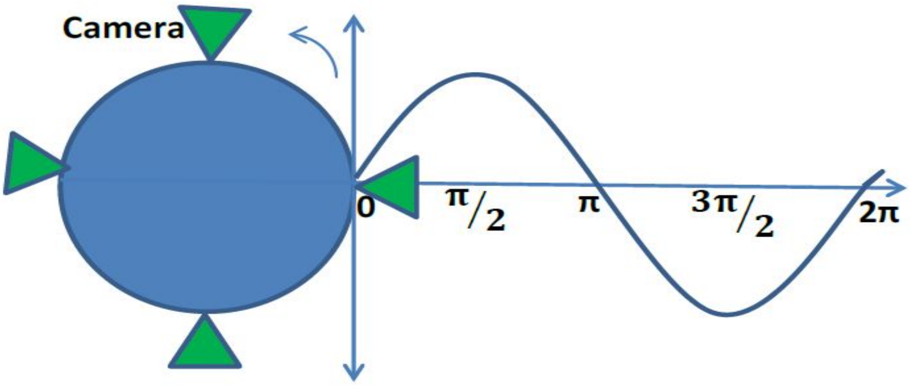


Figure 6.2: Relating Camera rotation to a sinusoid.

implementation challenges, to be able to use a set of cameras arranged on the perimeter of a trapezoid to capture a scene in multi-view video.

The other discussions in this chapter will unfold as follows. A comprehensive explanation of TCA will be given in section 6.3. In section 6.4, conceptual and implementation challenges are highlighted. Camera and depth map construction is the focus of section 6.5. Experimental results are presented in section 6.6. The conclusion of this work is drawn in section 6.7.

6.3 THE PROPOSED METHOD

Linear, divergence, and convergence camera arrangements have been used for acquiring visual content in multi-view video. One feature which is common to these camera configurations is that they are planar. All the cameras are constrained to the same, usually horizontal plane. We now explore the design space of trapezoid graph to seek the possibility of having some of the cameras positioned at a different vertical coordinate level.

The strong point of TCA is that it allows for adaptive camera topology by which points within the scene, especially the occluded ones can be optically and geometrically viewed from several different viewpoints either on the edge of the trapezoid or inside it.

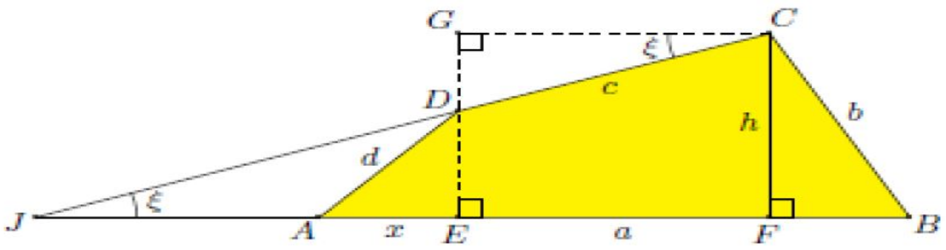


Figure 6.3: Alternate angle definition of quadrilateral.

6.3.1 *Trapezoid graph*

In a greater generality, in [330–334], a trapezoid can be obtained from a trapezoid graph, which is an undirected graph $G(V, E)$ consisting of a set of vertices $V = \{v_1, v_2, v_3, ..., v_n\}$ and a set of edges $E = \{e_1, e_2, e_3, ..., e_n\}$ if and only if $(v_i, v_j) \epsilon E$. A trapezoid computationally allows for the determination of a corner camera viewpoint and to be used in conjunction with the initially chosen corner viewpoint on the perimeter of the trapezoid. Using this concept, it is possible to determine any two cameras located on the corner points of the trapezoid that can be considered for stereoscopic analysis.

6.3.2 *Formal Mathematical Statement of TCA*

A trapezoid is considered based on the inclusive definition of having high significance in mathematical analysis. It is well known that any proven property of a trapezoid automatically holds for a rectangle, square, rhombic, parallelogram. Mathematically speaking, a quadrilateral in Figure 6.3 can be described using the relationship in (6.1).

$$(AC)^2 + (BD)^2 = (BC)^2 + (AD)^2 + 2(AC)(CD)\cos\xi. \tag{6.1}$$

Such a quadrilateral simplifies to a trapezoid when ξ equals zero as can be seen in Figure 6.3. This means CD is parallel to AC. This is the definition of a trapezoid that is adopted in this work.

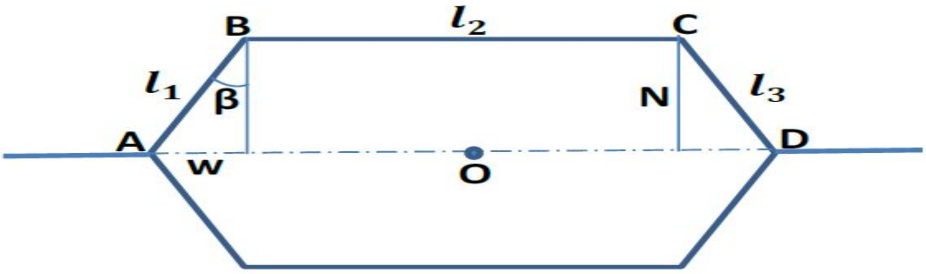


Figure 6.4: Trapezoidal representation.

Investigating the base width of a given trapezoid of height N shown in Figure 6.4 can provide useful information as to how the stereoscopic distance can be computed. The perimeter length, L of the upper-half hexagon is given as

$$L = l_2 + 2N\left(\frac{W^2}{N^2}+1\right)^{\frac{1}{2}}. \tag{6.2}$$

If the cross sectional area of the trapezoid is C_a, then

$$l_2 = \frac{C_a}{N} - N\left(\frac{W}{N}\right). \tag{6.3}$$

The N rate of change of L can be expressed as

$$\frac{\partial L}{\partial N} = -C_a N^{-2} + 2\left(\frac{W^2}{N^2}+1\right)^{\frac{1}{2}} + 2N\left(\frac{1}{2}\right)\left(\frac{W^2}{N^2}+1\right)^{-\frac{1}{2}}\left(-2\frac{W^2}{N^3}\right). \tag{6.4}$$

By taking maximum of (6.4) into consideration, the allowable cross sectional area, C_a of the trapezoidal can be obtained as in (6.5).

$$C_a = N^2\left[2\left(\frac{W^2}{N^2}+1\right)^{\frac{1}{2}} - \left(\frac{W^2}{N^2}+1\right)^{-\frac{1}{2}}\left(2\frac{W^2}{N^2}\right)\right]. \tag{6.5}$$

In a special case where W is small compared with N, C_a reduces to

$$C_a = 2N^2. \tag{6.6}$$

The expression in (6.6) implies that the area of the trapezoid is approximately twice the area of a square of side N.

When two cameras were positioned at A and D, the stereoscopic distance AD, is given by (6.7). The simplification of (6.7) indicates that distance AD equals twice parts l_1. This is a unique property of a trapezoid which expresses the distance between the bottom parts of the legs of a trapezoid as a function of trapezoid sizes.

$$AD = N\left[2\left(\frac{W^2}{N^2}+1\right)^{\frac{1}{2}} - \left(\frac{W^2}{N^2}+1\right)^{-\frac{1}{2}}\left(\frac{W^2}{N^2}\right)\right] + W. \qquad (6.7)$$

Therefore, the stereoscopic distance for any two cameras positioned along the base AD of the trapezoid can always be determined very easily.

Also the perimeter length L of the trapezoid is described as

$$L = 4N\left(\frac{W^2}{N^2}+1\right)^{\frac{1}{2}} - \left(\frac{W^2}{N^2}+1\right)^{-\frac{1}{2}}\left(\frac{W^2}{N^2}\right). \qquad (6.8)$$

Based on a similar analysis, the β rate of change of L simplifies to (6.9). It is clear from (6.9) that camera positioned on the corner points of a trapezoid can be made to have equal stereoscopic distance when the trapezoid is half-hexagon in which all the sides are equal. This observation provides a unique way to virtually increase the camera density.

$$\sin\beta = \frac{1}{2}. \qquad (6.9)$$

6.3.3 *Characterisation of Trapezoid*

A trapezoid has five important characteristics which can be explored in the computation of camera positions in TCA. These are sides and distances, collinearity, trigonometry, area and v-parallelogram characteristics [327].

6.3.3.1 *Sides and Distances*

In the trapezoid of Figure 6.4, triangles ACB and DBC have equal areas since they both have an altitude of N. Therefore, it can be written that

$$\frac{1}{2}(l_2)\left(\sin\left(\frac{\pi}{2}-\beta\right)\right)l_1 = \frac{1}{2}(l_2)\left(\sin\left(\frac{\pi}{2}-\beta\right)\right)l_3. \qquad (6.10)$$

The distance W can be obtained by equating the sum of the left and right hand-terms of (6.10) with (6.8). In this way, the coordinate description of the position of any camera located along AB can be easily obtained. Also, in the event of any cameras located at the point of intersection of diagonals AC and BD, the position coordinates of any such cameras can be determined based on the generalised Euler parallelogram law expressed in (6.11). X is the horizontal difference between the centre point of AC and vertex D, where $X = \frac{(AD-BC)}{2}$. In certain specialised applications, cameras positioned at the midpoints of the trapezoid diagonals can be explored to attempt to solve the Problem of occlusion.

$$(AB)^2 + (BC)^2 + (CD)^2 = (AC)^2 + (BD)^2 + 4X^2. \tag{6.11}$$

For any two cameras at centres of AB and CD, the stereoscopic distance SD, is computed as

$$SD = \frac{1}{2}\left(BC + AD\right). \tag{6.12}$$

Assuming two cameras were positioned at L and $L^{'}$ on the non-parallel sides AB and CD of the trapezoid of Figure 6.4. Let k be the same ratio by which L and L' divide the non-parallel sides. Applying the idea discussed in [229] and other relevant materials, stereoscopic distance $LL^{'}$ can be obtained as

$$LL^{'} = k\left(1-k\right)BC + kAD. \tag{6.13}$$

The dependency of the diagonals on the sides and β can be formulated for an efficient trapezoid (i.e all sides are equal) as

$$(AC)^2 = N^2 + (W + l_2)^2 \tag{6.14}$$

$$(BD)^2 = N^2 + (W + l_2)^2 \tag{6.15}$$

$$l_1^2 = l_3^2 = W^2 + N^2 \tag{6.16}$$

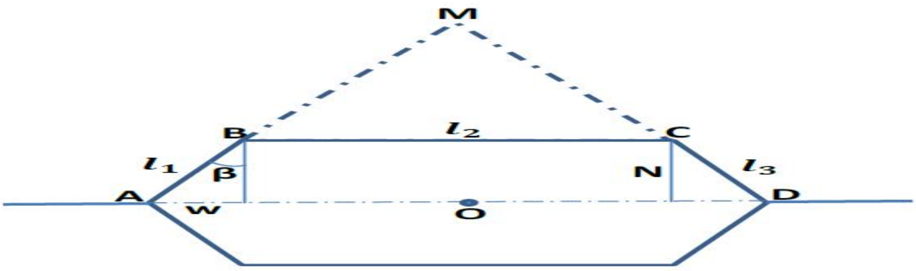


Figure 6.5: Trapezoid representation.

$$(AC)^2 + (BD)^2 = 4l_1^2\,(1 + \sin\beta)\,. \tag{6.17}$$

The relationship in (6.17) states that for an efficient trapezoid, the sum of the squares of the diagonal is directly proportional to the square of the side length l_1. The constant of proportionality equals 6. Again, when β is equal to zero degree, (6.17) simplifies to the expression expected for the sum of squares of the diagonal of a square. This is a further confirmation of the validity of (6.17).

In an attempt to solve the conceptual and implementation challenges, the trapezoid of Figure 6.5 will be used to define a scene. This will be further explained in section 6.4.

6.3.3.2 *Collinearity*

Consider a circle drawn through the corner points of the two halves of the trapezoids shown in Figure 6.4. By applying the analysis of Christopher Bradley [335], as demonstrated in Figure 6.6 to Figure 6.4, points similar to E, F, G, and N which are collinear will be obtained. This means that the point of intersection E of the diagonals of quadrilateral $ABCD$, the point of intersection G of the centroids of triangles BCD, ACD, ABD, ABC, the point of intersection N of the midpoints of AB, BC, CD, DA and the point of intersection F of the centroids of the triangles ABE, BCE, CDE, DAE, all lie on a line and hence are collinear. These points are potential candidate camera viewpoints which could be explored in an attempt to challenge the problem of point visibility.

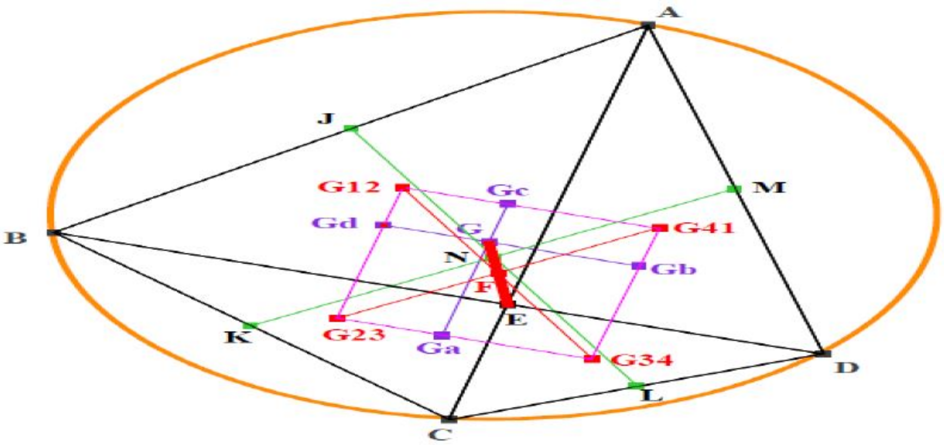


Figure 6.6: Cyclic quadrilateral.

6.3.3.3 *Area*

The area characteristic of a trapezoid is important in the computation of scene volume. An extension of the work in [258] provides a cue to the area A, of trapezoid as in

$$A = \frac{1}{4}\sin\vartheta\left(\left(2\left((AB)^2+(CD)^2\right)-4X^2\right)\left(2\left(BC^2+AD^2\right)-4X^2\right)\right)^{\frac{1}{2}}. \tag{6.18}$$

6.3.3.4 *Similarity*

Assuming in Figure 6.4 that the diagonals AC and BD intersect at a point P. Let the perpendicular lines from P to AD and BC have lengths h_1 and h_2 respectively. Also, θ_1 and θ_2 respectively be the angles substended by AC and BD with AD. The following two equations can be written

$$h_1 = AP\sin\theta_1 = PD\sin\theta_2 \tag{6.19}$$

$$h_2 = PC\sin\theta_1 = PB\sin\theta_2 \tag{6.20}$$

$$\frac{AP}{PC} = \frac{PD}{PB}. \tag{6.21}$$

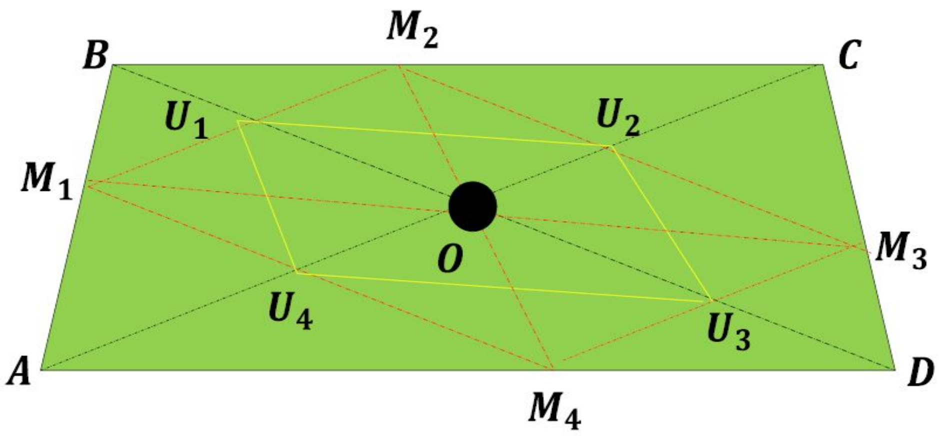


Figure 6.7: V-parallelogram M of a trapezoid.

Expression (6.21) is a trapezoidal similarity characteristic which can facilitate the computation of the coordinates of any cameras positioned at P.

6.3.3.5 *V-Parallelogram Characteristic*

V-parallelogram characterisation of trapezoids shown in Figure 6.7 has been at the centre of recent research investigation in mathematics aimed at solving problems of broad relevance [336, 337]. It is said to be formed when points on any adjacent sides of a trapezoid are joined with a line which is parallel to a diagonal of the trapezoid which does not go through the angle between the adjacent sides and the one opposite it. This definition implies that each corner point of a V-parallelogram lie on one and only one side of the trapezoid. Therefore, it means that cameras positioned on adjacent sides of a trapezoid can stereoscopically be used to observe a scene.

In multi-view visual content acquisition, cameras can also be positioned at viewpoints M_1, M_2, M_3, M_4 which , according to a statement in Euclidean geometry define a V-parallelogram $M_1M_2M_3M_4$ shown in Figure 6.7. The homothetic transformation of triangles ABC and M_1BM_2 around the corner point B of the trapezoid in Figure 6.7 provides for (6.22) from which the baseline distance can be determined.

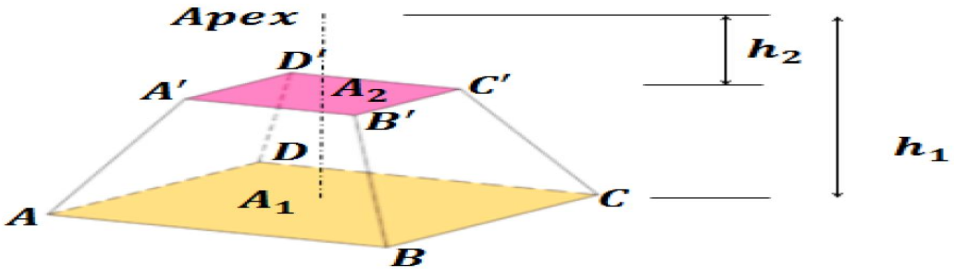


Figure 6.8: Implementation strategy of TCA.

$$\frac{AM_1}{AB} = \frac{CM_2}{BC} = \frac{AU_4}{AO}. \tag{6.22}$$

6.4 IMPLEMENTATION CONSIDERATION

Of course, TCA is potentially inclined to addressing some of the critical challenges in multi-view video acquisition; however its implementation could also be a challenge. Two issues easily come to mind namely the possible curvature of the scene and the frequency of the trapezoid. For a scene of large volume, the curvature of the opposite and parallel sides of the trapezoid becomes significant and adds up to the complexity of the implementation strategy. Practically, this situation implies that the stereoscopic distance can no longer be measured on a straight line between two camera positions.

Again, for a large scene where a small stereoscopic distance is required between cameras, several cycles of TCA will have to be implemented. This will significantly make the cost of implementing this strategy very high. A right balance will have to be struck between small and large scenes on one hand and curvature and frequency of TCA on the other.

To adequately address these issues, an implementation strategy is proposed based on the assumption that the scene is contained in a pyramidal frustum defined by four trapezoids as shown in Figure 6.8.

In this method, only four half-cycles of TCA are required for the entire scene to be adequately covered. The acquisition cameras are then placed on the edges of the trapezoid such that the ones along the legs of the trapezoid exhibit convergence topology. The idea of scene volume in this work does not require a strong argument for it to be embraced since it has been mentioned in [44] where a volumetric space is defined as an arbitrary connected 3D volumetric physical room which the arrangement of multiple cameras is expected to cover. The formulation of scene volume is actually a fortification of the basic volume (6.23) of a frustum. The modification of (6.23) will take into account the near and far depth parameters of the scene.

$$V = \frac{(h_2 - h_1)}{3}\left(A_1 + (A_1A_2)^{\frac{1}{2}} + A_2\right) \tag{6.23}$$

6.5 CONSTRUCTION OF CAMERA AND DEPTH MAP

Stereoscopic camera system with parallel axes is believed to be necessary to avoid the vertical image disparity generated by systems that verge the camera axes [181]. However, "Two forward-facing camera" technique [338] used to deliver a 2D projection of a scene remains a fundamental challenge in stereo vision. Correspondences between multiple images need to be accurately established. These are understood to be attributable, in the view of [338], to scene properties such as textureless areas, non-Lambertian surfaces, reflections and translucency, and occlusions. Camera issues include image noise, calibration and synchronisation, differences in exposure, white balancing, and other radiometric properties. A camera is formulated based on pin-hole camera model expressed in (6.24).

$$\Lambda x = KRT\begin{bmatrix} I| & -t \end{bmatrix} \tag{6.24}$$

K is a 3×3 matrix consisting of camera intrinsic parameters namely, focal length, aspect ratio, skew and principal point. R denotes an orthogonal matrix and T a vector, encoding camera rotation and translation respectively, depth, λ, is equal to Z. Both are determined by the orientation and displacement of the camera at the virtual viewpoint on the edge of the trapezoid. x and X_o are the extended image and scene point coordinates. The projection matrix P is used to render the quantisation level v, for each vertex and face of the graphic. The substitution of v into (6.24) yields the expected depth for all the vertices and faces. Using uniform quantisation of depth value between Z_{near} and Z_{far} to minimise dis-occlusion, and taken into consideration the fact that in humans, the perceived depth distance of close objects is much less than the depth distance of further objects [339, 340], depth is formulated as

$$depth = \frac{Z_{near}}{(1 - \frac{v}{2^{32}})} \tag{6.25}$$

where v is the number of quantisation levels from zero through to $2^{32} - 1$. The high number of quantisation levels is aimed at improving the resolution of the depth map.

6.6 EXPERIMENTAL RESULTS AND USER TRIAL

Some of the simulation results to validate the concept of TCA are shown in Figure 6.9. More results specific to each object are shown in Figure 6.10, Figure 6.11, Figure 6.12, Figure 6.13, and Figure 6.14. Each disparity map of an object is generated using a different virtual viewpoint. These are high quality disparity map. In the analysis leading to the determination of disparity map, correspondence problem still remain a major source of error. This is because correlation based techniques are widely used. It is formulated based on resemblance constraint which requires intensity similarity between any corresponding points [338]. Also, it cannot cope with texture-less regions where pixels are insufficiently distinct. Occlusions and discontinuities of some parts of scene are important features which correlation methods cannot adequately handle. The choice of window size is critical with regards to accurate disparity map.

To reduce the error contribution due to the aforementioned, comparison of quantitative results with normalised cross-correlation method, Daisy descriptor, error quadratic mean, and Local evidence have been employed in most research work related to depth map which are documented in literature. Since only a single image of the different objects is used in this work, i.e not stereo images, the earlier mentioned objective quality metrics cannot be applied. Hence subjective assessment is conducted as explained in the next three paragraphs.

However, Subjective Quality Assessment (SQA) of the generated disparity maps has been conducted at Centre for Media Communication Research (CMCR), Brunel University, London. Five groups of depth maps images were chosen for consideration. These are Chair, Teapot, Space-shuttle, X29-plane, and Canstick.

Of the several ITU-R recommendation BT.500 methods, Double Stimulus Impairment Scale (DSIS) is adopted. In general, the subjective rating based on double stimulus method has the advantage of not being susceptible to the severity and ordering of the impairments within the test session [341]. This method involves scoring the change in quality between a reference frame and target (impaired) frame after separate observation.

To implement this method, for each group of depth maps, twenty research students between the age of 28 and 40, who already have a good understanding of disparity map were invited one after the other to take a look at the disparity maps for different objects and give their judgement of the quality of the depth maps. The disparity maps were displayed on a computer screen and observed using the nicked eyes.

The analysis leading to SQA was based on DSIS in which the following scales were used. “Very clean and sharp = 5”, “Clear but not sharp = 4”, “Clear but blur = 3”, “blur = 2”, and “Not clear = 1”. The calculated mean opinion scores (MOSs) for each disparity map were then analysed. More than 90 percent of the observers scored each disparity map with DSIS of 5.

In another important experiment performed in Maya software environment to validate the concept of TCA, a static scene is created. The scene is assumed to be contained in a frustum as depicted in Figure 6.8. On each of the four sides of the scene is arranged four auto-focus cameras which form a trapezoid graph. A snap shot of the scene from some of the cameras are shown in Figure 6.15 as perspective views.

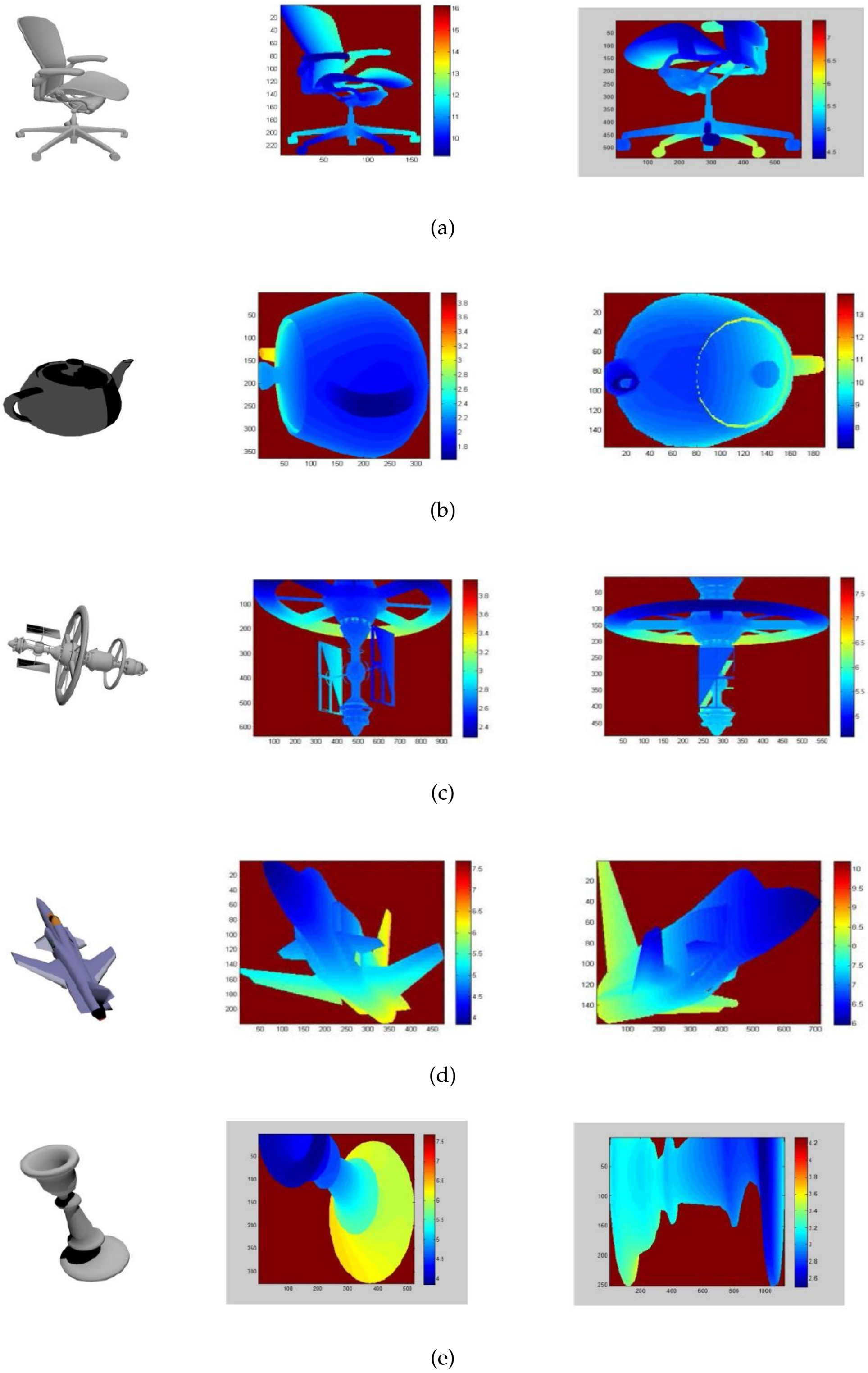
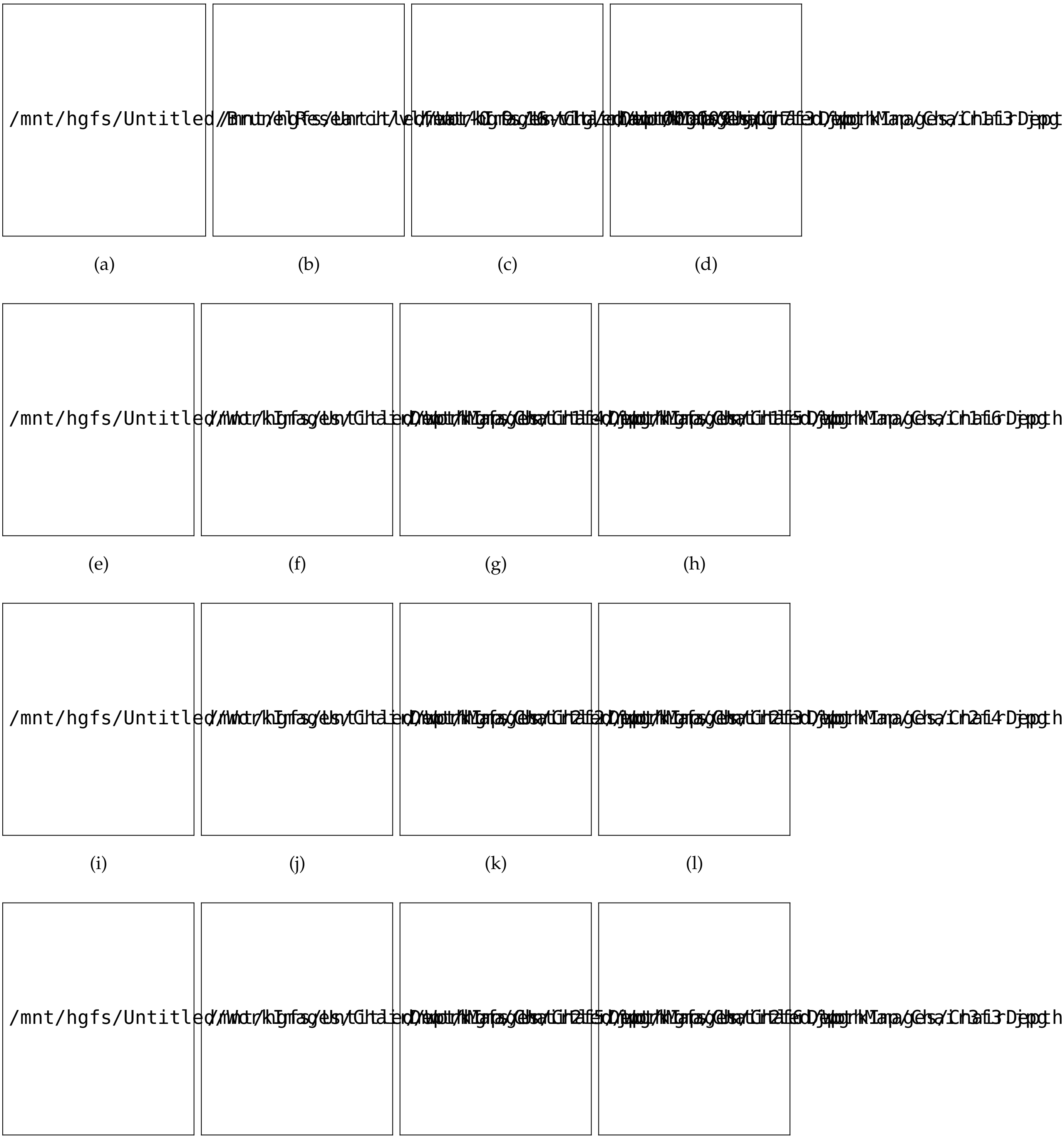


Figure 6.9: Different objects are shown in the first column. The disparity maps for different views of each object are shown on the same row.



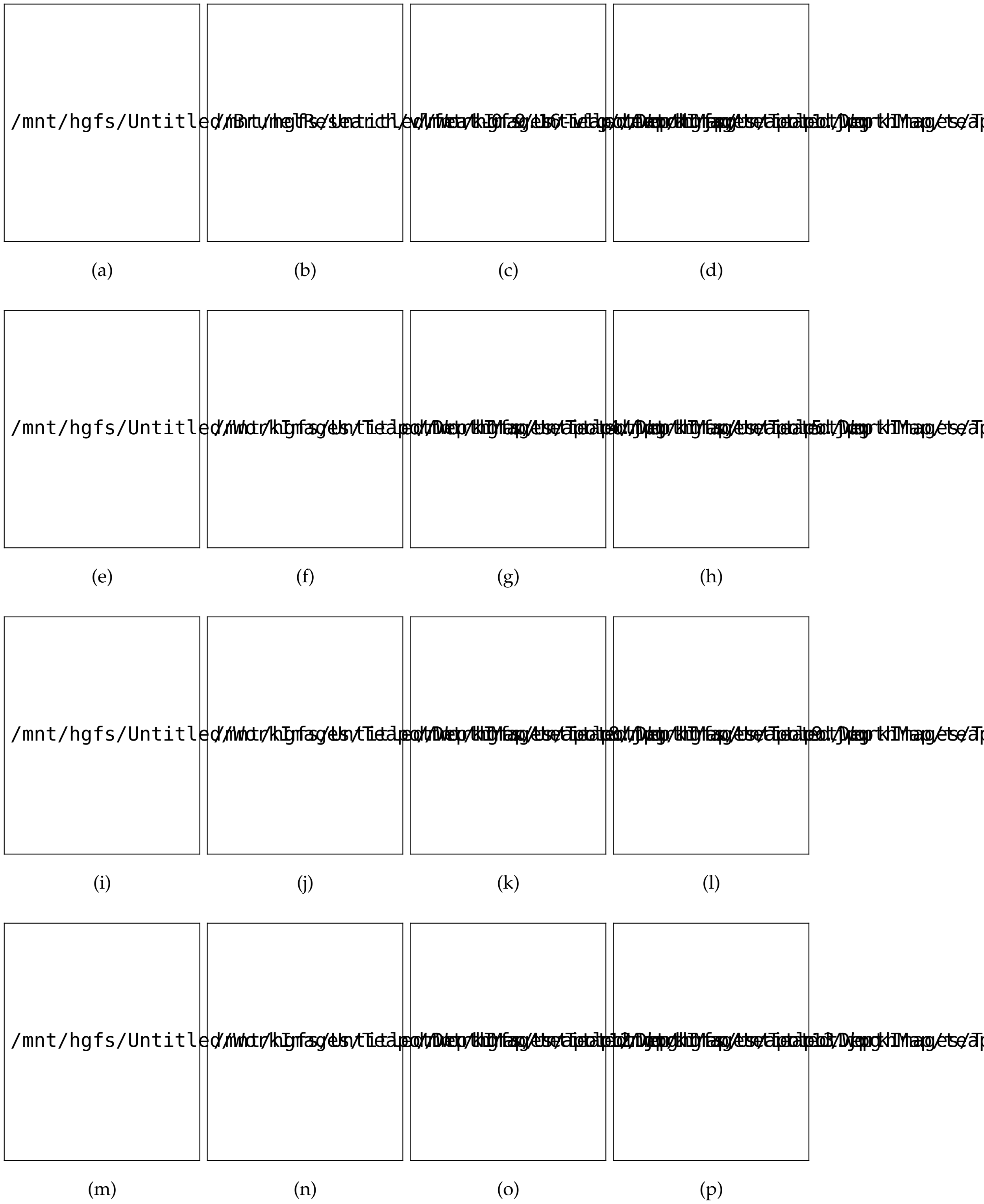
(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Figure 6.10: A chair is shown in (a). The disparity maps for different views of the chair are in (b) through to (p).



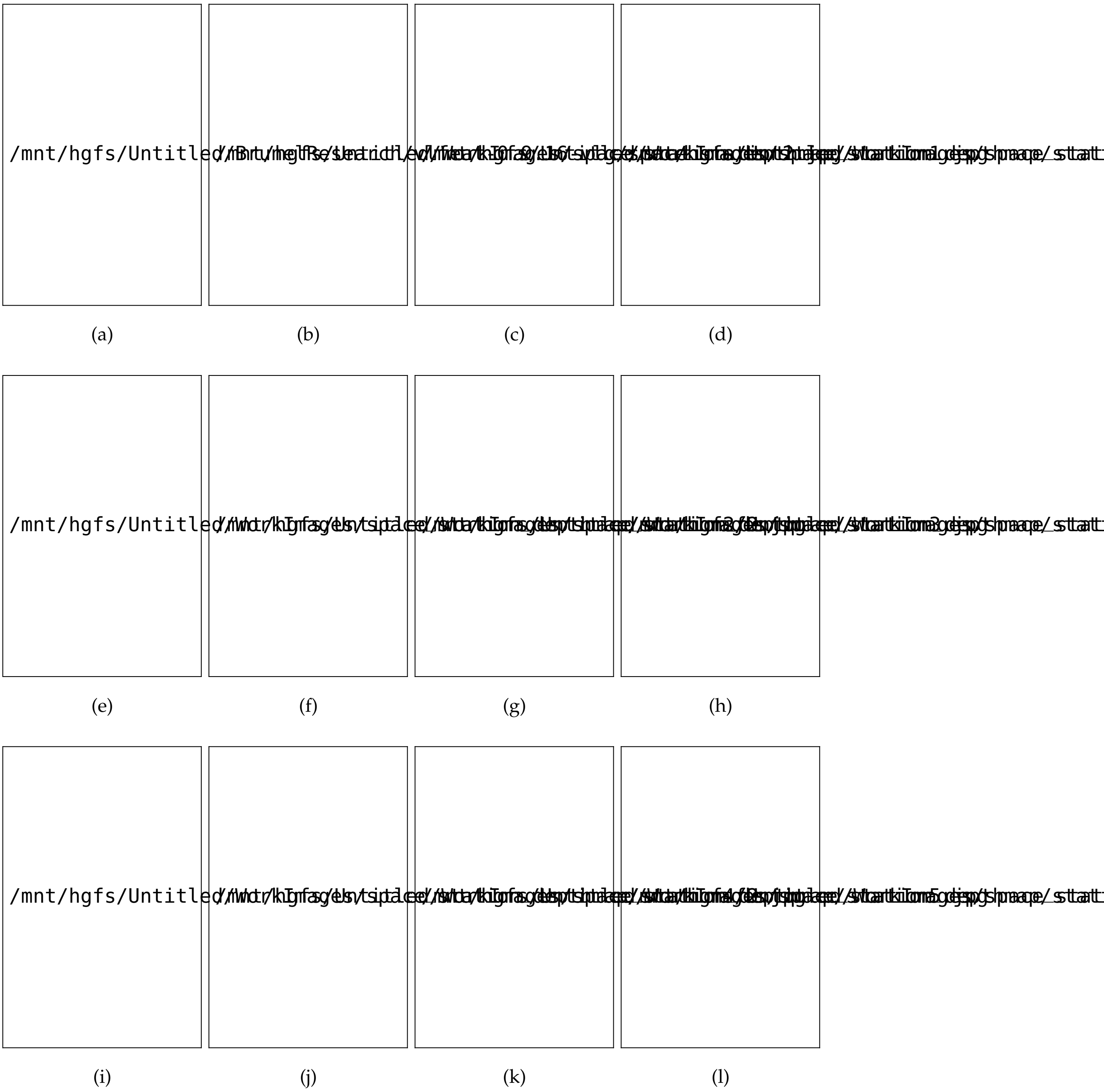
(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Figure 6.11: A teapot is shown in (a). The disparity maps for different views of the chair are in (b) through to (p).

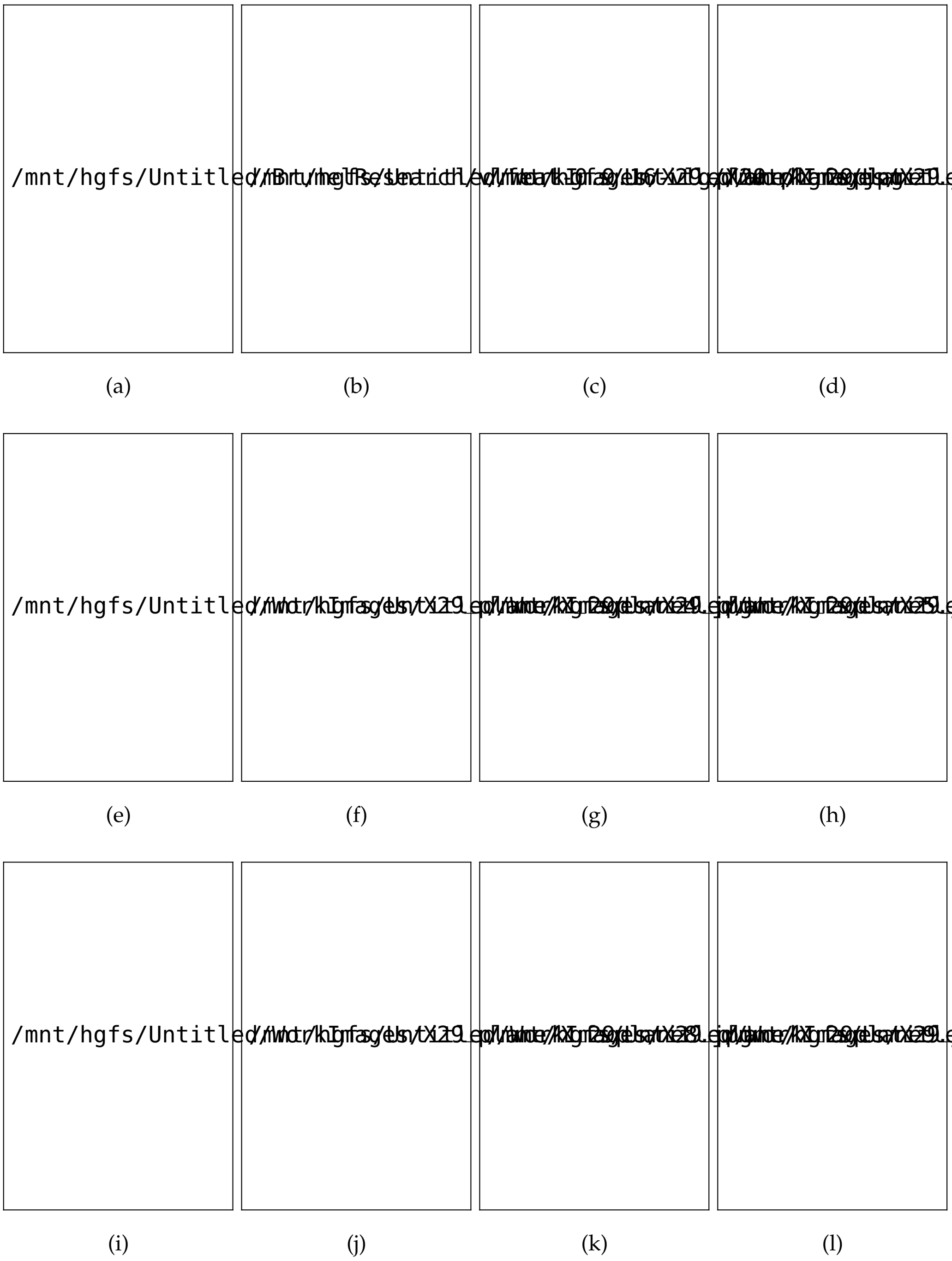


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 6.12: A space-station is shown in (a). The disparity maps for different views of the chair are in (b) through to (p).

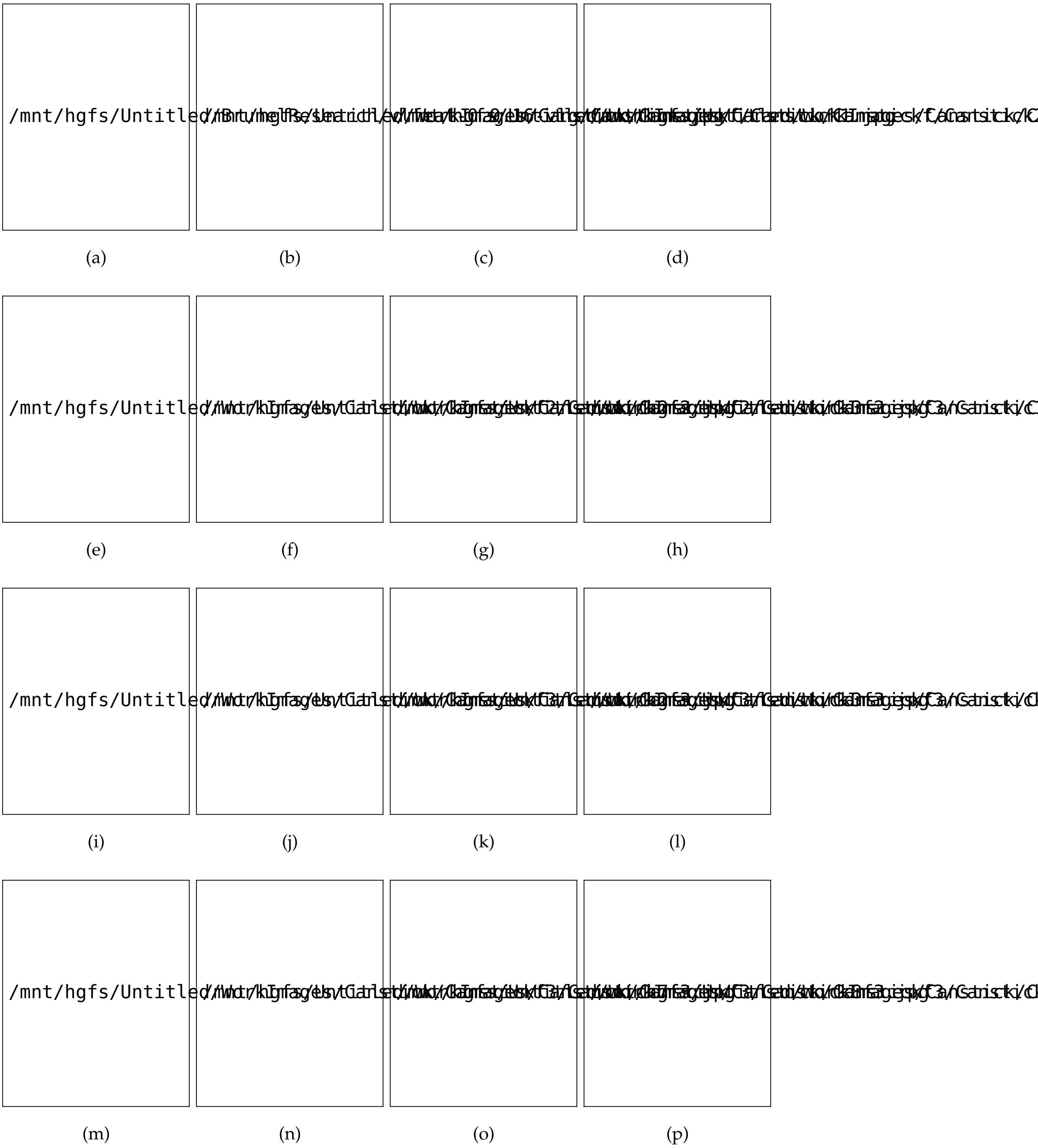


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 6.13: A X29-plane is shown in (a). The disparity maps for different views of the chair are in (b) through to (p).



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

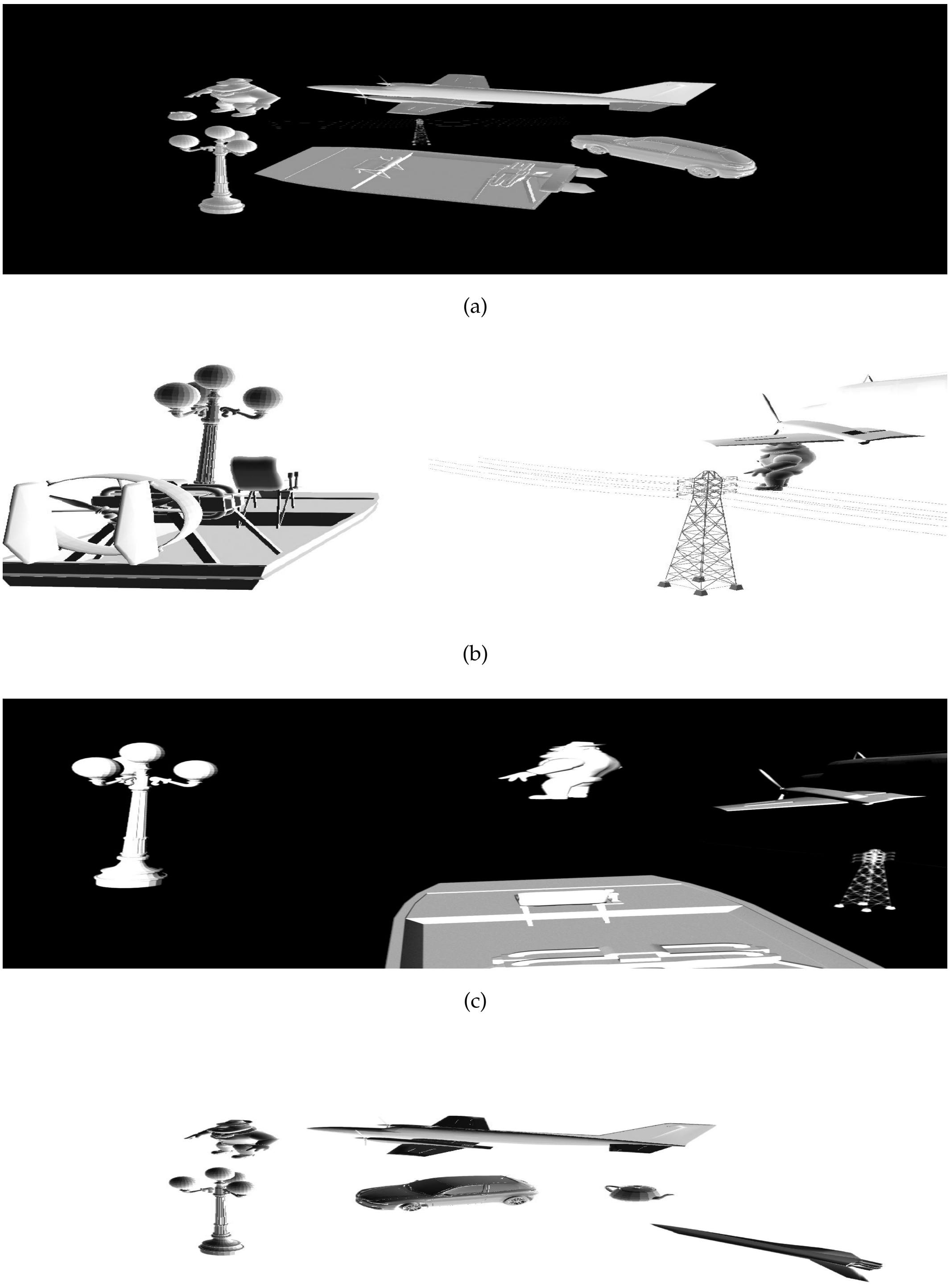
(m) (n) (o) (p)

Figure 6.14: A canstick is shown in (a). The disparity maps for different views of the chair are in (b) through to (p).

Another dimension of the experiment in Maya is the creation of Gray scale depth map. This means that the scene depth levels are expressed in terms of transition from black to white colours. Gray scale depth maps for different objects are shown in Figure 6.16 As the scene depth increases, the intensity of light decreases from white: for points closest to the camera, to black which represent the background of the scene.

6.7 CHAPTER SUMMARY

The description of TCA has been presented based on an efficient trapezoid. In principle, the definition of a trapezoid is simple. However, it is the diverse different methods of determining the coordinates of its corner points, points on the edges, and in some critical applications, points in the space defined by the edges that make TCA so appealing in camera positioning for multi-view video. The similarity, sides and distances, areas, and trigonometric characteristics of a trapezoid provide for computation of baseline and coordinate description of camera positions. Also the understanding provided by TCA has been used for depth map generation with respect to virtual viewpoints on the edge of the trapezoid. Virtual viewpoints which trace a trapezoid guarantee high quality disparity map as indicated by DSIS, with almost complete absence of holes. Also, the idea of scene visibility and depth map has been demonstrated in Maya were different perspective view of the scene have been show with respect to viewpoints which trace a trapezoid graph.



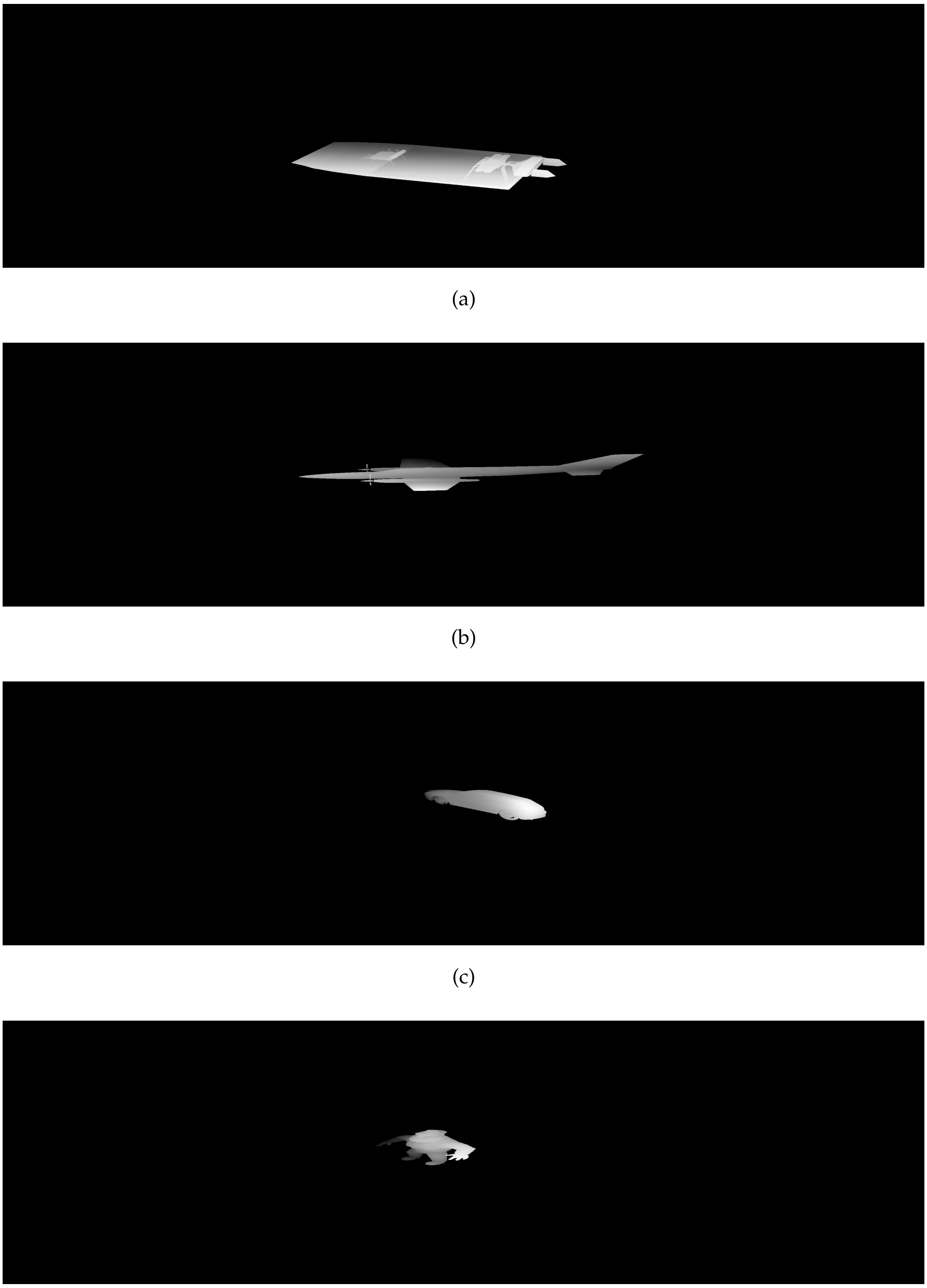
(a)

(b)

(c)

(d)

Figure 6.15: Different perspectives of a scene rendered in Maya software environment based on TCA. The perspectives represented in (a) through to (d) were rendered from different viewpoints on the edge of a trapezoid.



(a)

(b)

(c)

(d)

Figure 6.16: Depth map rendering of some scene objects in Maya software environment based on TCA. In the depth maps in (a) through (d), the closest part of the scene object to the camera is in white colour. This gradually fades away towards the furthest point from the camera which is represented in black.

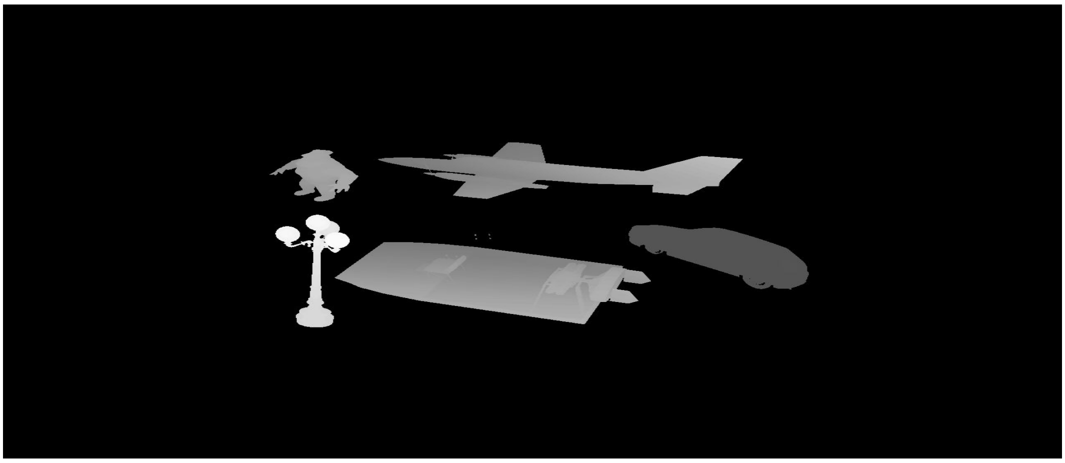


Figure 6.17: Depth map rendering of a scene consisting of several individual objects in Maya software environment based on TCA. The closest part of the scene object to the camera is in white colour. This gradually fades away towards the furthest point from the camera which is represented in black.

Part IV

CONCLUSION

7

SUMMARY AND OUTLOOK

7.1 INTRODUCTION

In this chapter is a focused mention of what has been achieved in this research. This thesis was aimed to realise virtual camera and trapezoidal camera architecture. Also, a straightforward conclusion and future direction of the research contributions discussed in this thesis are highlighted.

7.2 OVERVIEW

The acquisition of visual content has been the main focus of this thesis. Usually, in multi-view video, visual content is obtained by using multiple real cameras. The formation of a network of real cameras is fundamentally based on either parallel, convergence, or divergence topology. Irrespective of the topology used, there is always a functional relationship between the quality of the created visual content and camera density. Dense real camera architecture has been proven to be very effective despite the cost estimate, geometrical calibration, colour balance between the individual cameras, mechanical limitation, and temporal synchronisation issues. The ability to reconstruct scene objects from a series of image samples is extremely valuable in a range of applications as it potentially allows diagnostic technology with superior sensitivity and selectivity.

One aspect of the research contribution presented in this thesis is the realisation of virtual cameras by using the method of IBR. A virtual camera is usually realised in software by the interpolation of pixels information contained in the images acquired by physical cameras. Two main variations of IBR have been used. The first being rendering with explicit geometry and the second method being rendering with no geometry. IBR method which uses explicit geometry has been applied in the study of the effect of region match measures on rendered images.

The considered metrics are SAD, SSD, SCP, and NSCP. The reason for this step is that the problem of finding the pixel coordinates in two different intensity images that correspond to the same point in the world is recognised as a data association problem. This deserves that stringent measures are used in the detection, extraction, and matching of interest points. This information was then used in the construction of a depth map. Morphological operations based on a suitable structuring element was then used to fill any resulting holes associated with the constructed depth map. A 3D model of the scene is created by depth map texturing. The positive consequence of this is the rendering of high quality image with respect to a virtual viewpoint by re-projection process.

The no geometry technique of IBR has been thoroughly investigated and used to construct a 360 degree panoramic image characterised with depth perception. Two categories of 3D panoramic content have been demonstrated. 3D content from multiple images captured by a single panned camera and the use of stereo-cameras on a single tripod. In both categories, the implementation is at two broad levels. First, the methods of feature detection, extraction, and matching was further explored in the computation of homography and motion model which facilitates the stitching of a pair of images. With regards to the generation of 3D from multiple image acquired using a single panned camera, a virtual panoramic image is also generated. Therefore, if the real panoramic image is assumed to be the left image, then the virtual panoramic image constitute the right one. Second, anaglyph process which involves the superposition of two panoramic images (left and right image) is performed to generate 3D effect. It contains a significant number of effects designed to improve realism, attractiveness of attention, considerable emotional response. Theses effects are depth, viewing discomfort, quality, naturalness, immersive feeling. All these manifest into more visual information within a given time frame through the use of anaglyph glass.

The generation of 3D effects from two panoramic images has a special fascination. However, panoramic 3D effect has also been created from a set of images captured using a single camera. Using the real multiple images, a real 360 degree panorama is created. From this, another virtual panorama is created. Therefore, the real panorama constitute the left panorama and the virtual panorama forms the right panorama. The left and right panoramas were then superimposed to create depth perception. Different viewing depths have been generated and observed through a pair of anaglyph glasses with a good viewing experience.

The third part of the use of IBR in the realisation of virtual camera presented in this thesis is the computation of virtual environment from stereo panoramic image. Here both IBR without and with geometry have been used. IBR without geometry is used in the stitching of multiple images to form a pair of panoramic images. The creation of 3D model from depth map is achieved by using IBR with geometry. The depth map is constructed based on the normalisation of matching points and the singular value decomposition of fundamental matrix. A new image is then rendered with respect to a virtual viewpoint.

Another dimension of IBR that has been considered in this research is the reconstruction of 3D scene from multiple image samples with the same centre of projection. Three scenarios have been considered. One case considered the use of corrupted randomly generated pixel points of at a particular resolution based on a virtual camera of known intrinsic parameters. An attempt is then made to recover the corresponding scene. In the second situation, starting with real images, classic bundle adjustment algorithm from computational vision is used to recover the real scene related to the different optical and geometric properties contained in the image samples. The third case encompasses the use of real scene image samples, however, advantage is taken of the existence of sparsity. This means there is no interaction between the parameters for different 3D points and cameras and hence the linearised problem formulation has many zeros depicted as sparse block structure. This considerably helps to reduce the computational complexity of the algorithm. To achieve this, the C/C++ program at the heart of SBA algorithm is modified and access to the optimised parameters is made possible. Also the performance of SBA is considerable and can be gauged from the statistical summary presented in appendix section. However, further improvement is required in the visualisation technique that can cope with images with SCOP. This will enhance the quality of the rendered images.

Beyond this metric reconstruction point, another aspect of the research contribution presented in this thesis is the proposition of trapezoidal camera architecture TCA. TCA is based on an efficient trapezoid in which all sides are equal. Both the theoretical analysis to describe the potential of TCA and its implementation results have allowed for the potential of trapezoid graph to be appreciated and explored in an attempt to address challenging issues associated with camera structures in multi-view video. With TCA, some sets of any two cameras could differ in both x and y coordinates. This means not all the cameras are constrained to the same horizontal plane. This characteristic exhibited by TCA has been employed to demonstrate how a scene point could be viewed from different points. Disparity maps of scene objects with respect to viewpoints which trace a trapezoid have been constructed. As part of the experiments employed to evaluate the strength of TCA, a scene was created and captured in Maya. A network of cameras forming a trapezoid on each of the four sides of the scene is used to do the capturing which provide different perspective views of the scene. The evaluation of TCA in the Maya environment is experimentally relevant since it also allows the use in simulation of cameras of different intrinsic parameters (in particular focal length).

Where possible, an objective subjective assessment of some of the resultant images have been given based on MTF. From literature, MTF is considered to be the most important image quality factor. It is basically a measure of contrast. An objective assessment using MTF presents a specific advantage in that it is considered to be a measure of sharpness, the most important photographic image quality factor. The definition and description of MTF has been discussed in the appendix section. SQF which is measure of perceived image quality has also been used in other situations. It is only indirectly related to the perceived sharpness when an image is viewed.

7.3 LIMITATION OF PRESENT WORK

As part of the limitations associated with this research, there is an emphasis on the use of static scene to better elucidate virtual camera realisation, metric reconstruction, and TCA concept. The proposed TCA has been implemented with respect to static scenes. But in reality, every practical application of "structure from motion" algorithms constantly requires dealing with critical motions and dynamic scenes. Another dimension of the challenges that have not been considered in this work is the use of different camera models. In any advanced visual acquisition set-up, it should be possible to use cameras with different focal length and from different manufacturers.

7.4 FUTURE WORK

The practical implementation of TCA concept presented in this thesis amounts to a new set of challenges. Therefore, future work is necessary. A further area which requires research attention in visual content acquisition is the implementation of TCA for a unavoidable dynamic scenes. This is necessary since in visual content acquisition outside computer graphics procedure, dynamic scene constitutes a significant proportion of the scenes encounter in practice. In such a scenario, a comprehensive understanding of the computational complexity of TCA becomes crucial. Any positive development in this direction will not only provide a judgement of the significance of TCA but further commend it for real-time applications. All these indicate the need for a systematic analysis.

7.5 FINAL REMARK

In view of the above research methods and considerations, virtual camera has been brought into a central challenge as a visual content acquisition strategy outside computer graphics procedure. It has been sufficiently proven that virtual cameras can provide for excellent minimisation of physical cameras through pixel interpolation based on IBR and that TCA is a new dimension of multi-view video camera configuration which allows a scene object of interest to be viewed from several reliable viewpoints which trace a trapezoid graph.

Part V

APPENDIX

A

APPENDIX

Some discussions related to image quality assessment are presented in this section. Also, it contains some statistics related to SBA.

A.1 QUALITY OF EXPERIENCE

The quality of images are known to be affected by the camera sensor and lens, and image processing technique employed. Generally, image quality and quality in general usually finds itself subject to human judgement [186]. Quality of experience **QoE!** can be measured either subjectively or objectively. Unfortunately, subjective assessments are problematic because human vision is variable. While one observer may be able to perceive differences between 40 pairs of alternating lines in a millimetre, another may only distinguish 20 at the same viewing distance. To complicate matters, an individual's perception can vary by as much as 10% at different times.

However, the objective quantification of 2D image quality can be achieved in several different ways. The popular criteria are sharpness, noise, lens distortion, lateral chromatic aberration, texture detail, colour accuracy, blemishes, colour, software artefacts etc [342]. Sharpness which is a measure of spatial frequency response is the most effective. It determines the amount of detail an imaging or processing system can reproduce. It is also referred to as modulation transfer function MTF. MTF is one of the most important parameters by which image quality is measured. This function indicates, objectively, the "resolving potential" of the lens or processing stage. Optical designers and engineers frequently refer to MTF data, especially in applications where success or failure is contingent on how accurately a particular object is imaged. To truly grasp MTF, it is necessary to first understand the ideas of resolution and contrast, as well as how an object's image is transferred from object to image plane.

MTF is said to be a better criterion since it applies to the shape of the entire spatial frequency range. Hence it is synonymous in practice with Spatial Frequency Response (SFR). MTF is one of the best tools available to quantify the overall imaging performance of a system in terms of resolution and contrast. and is expressed as

$$MTF = \frac{contrast\, of\, output\, image}{contrast\, of\, input\, image} \tag{A.1}$$

The contrast m of an image is described as

$$m = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{A.2}$$

where I_{max} and I_{min} are the maximum and minimum pixel intensity respectively in the output and input images.

"A parameter that does carry information about the signal contrast is the MTF, which may be used as a measure of sharpness. The MTF describes the amplitude or relevant contrast by which sine functions of different frequencies are moderated by an imaging system. A MTF value of 1 indicates that the full amplitude is transferred by the imaging system, while a MTF value of 0 indicates that no signal at all is transferred. One dimensional MTF may be calculated from the Line Spread Function (LSF) as the absolute value of the Fourier transform of the LSF, when the LSF is known. The definition of the MTF presupposes that data is measure continuously" [343].

With reference to a specific spatial frequency, if V_B is the minimum luminance (or pixel value) for black areas— at low spatial frequencies (the frequency should be low enough so that contrast does not change if it is reduced), V_W is the maximum luminance for white areas— at low spatial frequencies, V_{min} is the minimum luminance for a pattern near spatial frequency f (a "valley" or "negative peak"), and V_{max} is the maximum luminance for a pattern near spatial frequency f (a "peak"). $C(0)$ in (A.3) is the low frequency (black-white) contrast.

$$C(0) = \frac{V_W - V_B}{V_W - V_B} \tag{A.3}$$

$C(f)$ in (A.4) is the contrast at spatial frequency f. Normalizing contrast in this way— dividing by $V_{max} + V_{min}$ ($V_{max} + V_{min}$ at low spatial frequencies)— minimizes errors due to gamma-related non-linearities in acquiring the pattern.

$$C(f) = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \tag{A.4}$$

Therefore the MTF at a given spatial frequency is computed as

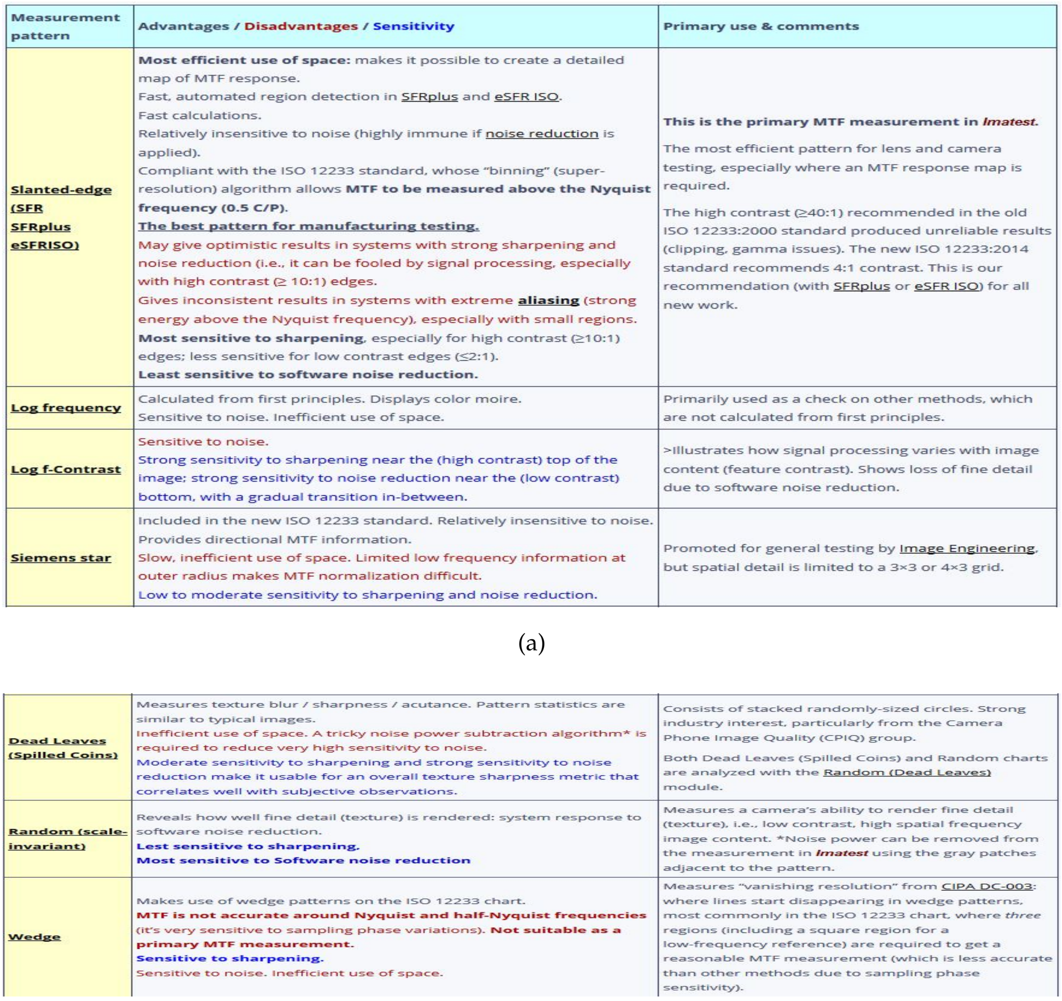
$$MTF(f) = \frac{C(f)}{C(0)} * 100\% \tag{A.5}$$

However, Stereoscopic **QoE!** is largely influenced by scene content, camera baseline, screen size and viewing position [344]. Many other ways of measuring MTF are shown in Figure A.1.

A.2 RESOLUTION TEST TARGETS

Test Targets are used to measure the accuracy or performance of an imaging system to ensure effective functioning. Test Targets are selections of slides or windows that are integrated into imaging systems to measure a number of imaging characteristics including resolution, distortion, or colour/grey scale. Test Targets often feature an array of lines, dots, or other patterns which an imaging system focuses on in order to determine its level of precision. Test Targets allow imaging systems to maintain a high level of accuracy over time or multiple applications.

Edmund Optics offers a wide range of Test Targets suited for many imaging systems. Test Targets are designed to easily integrate into an imaging system for quick, easy set-up. Test Targets may be used repeatedly with any number of different systems, allowing a small number of targets to service large numbers of imaging systems. Test Targets are available in glass, chrome, or photographic paper for multiple compatibility options for ease of use. Examples of test targets from Edmund optics are shown in Figure A.2 and Figure A.3.



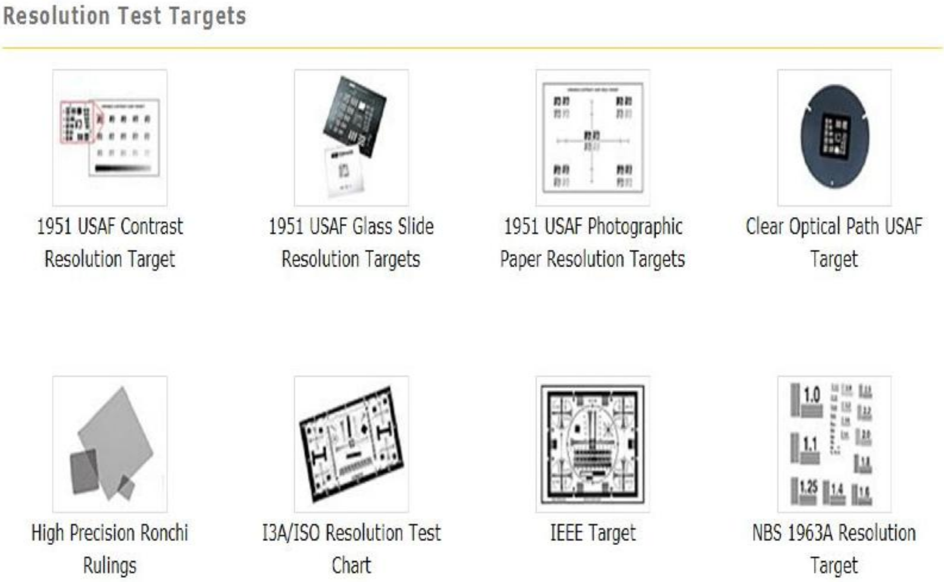
Measurement pattern	Advantages / Disadvantages / Sensitivity	Primary use & comments
Slanted-edge (SFR SFRplus eSFRISO)	**Most efficient use of space:** makes it possible to create a detailed map of MTF response. Fast, automated region detection in SFRplus and eSFR ISO. Fast calculations. Relatively insensitive to noise (highly immune if noise reduction is applied). Compliant with the ISO 12233 standard, whose "binning" (super-resolution) algorithm allows **MTF to be measured above the Nyquist frequency (0.5 C/P).** **The best pattern for manufacturing testing.** May give optimistic results in systems with strong sharpening and noise reduction (i.e., it can be fooled by signal processing, especially with high contrast (≥ 10:1) edges. Gives inconsistent results in systems with extreme **aliasing** (strong energy above the Nyquist frequency), especially with small regions. **Most sensitive to sharpening**, especially for high contrast (≥10:1) edges; less sensitive for low contrast edges (≤2:1). **Least sensitive to software noise reduction.**	**This is the primary MTF measurement in *Imatest*.** The most efficient pattern for lens and camera testing, especially where an MTF response map is required. The high contrast (≥40:1) recommended in the old ISO 12233:2000 standard produced unreliable results (clipping, gamma issues). The new ISO 12233:2014 standard recommends 4:1 contrast. This is our recommendation (with SFRplus or eSFR ISO) for all new work.
Log frequency	Calculated from first principles. Displays color moire. Sensitive to noise. Inefficient use of space.	Primarily used as a check on other methods, which are not calculated from first principles.
Log f-Contrast	Sensitive to noise. Strong sensitivity to sharpening near the (high contrast) top of the image; strong sensitivity to noise reduction near the (low contrast) bottom, with a gradual transition in-between.	>Illustrates how signal processing varies with image content (feature contrast). Shows loss of fine detail due to software noise reduction.
Siemens star	Included in the new ISO 12233 standard. Relatively insensitive to noise. Provides directional MTF information. Slow, inefficient use of space. Limited low frequency information at outer radius makes MTF normalization difficult. Low to moderate sensitivity to sharpening and noise reduction.	Promoted for general testing by Image Engineering, but spatial detail is limited to a 3×3 or 4×3 grid.

(a)

Dead Leaves (Spilled Coins)	Measures texture blur / sharpness / acutance. Pattern statistics are similar to typical images. Inefficient use of space. A tricky noise power subtraction algorithm* is required to reduce very high sensitivity to noise. Moderate sensitivity to sharpening and strong sensitivity to noise reduction make it usable for an overall texture sharpness metric that correlates well with subjective observations.	Consists of stacked randomly-sized circles. Strong industry interest, particularly from the Camera Phone Image Quality (CPIQ) group. Both Dead Leaves (Spilled Coins) and Random charts are analyzed with the Random (Dead Leaves) module.
Random (scale-invariant)	Reveals how well fine detail (texture) is rendered: system response to software noise reduction. **Lest sensitive to sharpening.** **Most sensitive to Software noise reduction**	Measures a camera's ability to render fine detail (texture), i.e., low contrast, high spatial frequency image content. *Noise power can be removed from the measurement in ***Imatest*** using the gray patches adjacent to the pattern.
Wedge	Makes use of wedge patterns on the ISO 12233 chart. **MTF is not accurate around Nyquist and half-Nyquist frequencies** (it's very sensitive to sampling phase variations). **Not suitable as a primary MTF measurement.** **Sensitive to sharpening.** Sensitive to noise. Inefficient use of space.	Measures "vanishing resolution" from CIPA DC-003: where lines start disappearing in wedge patterns, most commonly in the ISO 12233 chart, where *three* regions (including a square region for a low-frequency reference) are required to get a reasonable MTF measurement (which is less accurate than other methods due to sampling phase sensitivity).

(b)

Figure A.1: Ways to measure MTF. (b) is a continuation of (a).



Resolution Test Targets

1951 USAF Contrast Resolution Target

1951 USAF Glass Slide Resolution Targets

1951 USAF Photographic Paper Resolution Targets

Clear Optical Path USAF Target

High Precision Ronchi Rulings

I3A/ISO Resolution Test Chart

IEEE Target

NBS 1963A Resolution Target

Figure A.2: Test target.

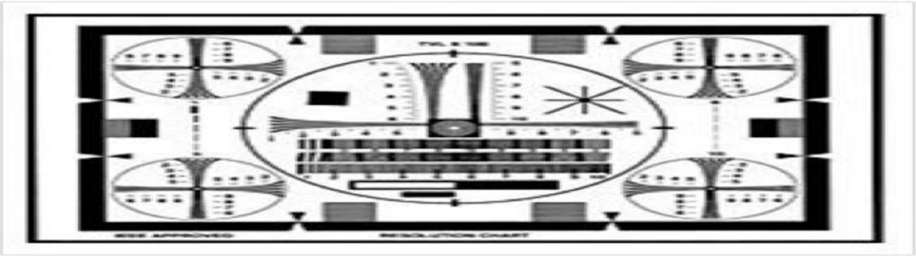


Figure A.3: IEEE test target.

A.3 SUMMARY OF MINIMISATION FOR IMAGE PAIR DSC_6731A-DSC_6732A

SBA using 464 3D pts, 2 frames and 928 image projections, 1404 variables.

Method BA_MOTSTRUCT, expert driver, analytic Jacobian, fixed intrinsics, without covariances.

SBA returned 150 in 150 iter, reason 3, error 0.0108958 [initial 1235.07], 250/150 func/fjac evals, 249 lin. systems.

Elapsed time: 0.20 seconds, 200.00 msecs

BIBLIOGRAPHY

[1] K.-H. Yap, L. Guan, S. W. Perry, and H. S. Wong, *Adaptive image processing: a computational intelligence perspective*. Crc Press, 2010. (Cited on page 3.)

[2] M. Tanimoto, "Ftv: Free-viewpoint television," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 555–570, 2012. (Cited on pages 3, 30, 41, 52, and 73.)

[3] A. J. Parker, "Binocular depth perception and the cerebral cortex," *Nature reviews.Neuroscience*, vol. 8, no. 5, pp. 379–391, 2007. (Cited on page 4.)

[4] B. Mendiburu, *3D movie making: stereoscopic digital cinema from script to screen*. CRC Press, 2012. (Cited on page 4.)

[5] D. Rosenthal, P. DeGuzman, L. C. Parra, and P. Sajda, "Evoked neural responses to events in video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 358–365, 2014. (Cited on page 4.)

[6] N. Brosch, A. Hosni, G. Ramachandran, L. He, and M. Gelautz, "Content generation for 3d video/tv," *e and i Elektrotechnik und Informationstechnik*, vol. 128, no. 10, pp. 359–365, 2011. (Cited on page 4.)

[7] S. Ince and J. Konrad, "Occlusion-aware view interpolation," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–15, 2008 2008. M3: Article. (Cited on page 5.)

[8] K. MÃŒller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3d video systems," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–11, 2008 2008. M3: Article.

[9] M. Lambooij, K. Hinnen, and C. Varekamp, "Emulating autostereoscopic lenticular designs," *IEEE/OSA Journal of Display Technology*, vol. 8, no. 5, pp. 283–290, 2012.

[10] J. Luo, K. Qin, Y. Zhou, M. Mao, and R. Li, "Gpu rendering for tiled multi-projector autostereoscopic display based on chromium," *Visual Computer*, vol. 26, pp. 457–465, 06 2010. M3: Article. (Cited on page 5.)

[11] L. Onural, *3D video technologies : an overview of research trends*. Bellingham, Wash: SPIE, c2010. (Cited on pages 5, 30, and 73.)

[12] Y. fan Peng, H. feng Li, Q. Zhong, X. xing Xia, and X. Liu, "Large-sized light field three-dimensional display using multi-projectors and directional diffuser," *Optical Engineering*, vol. 52, no. 1, pp. 017402–017402, 2013. (Cited on page 5.)

[13] Y. Peng, H. Li, Q. Zhong, and X. Liu, "59.2: Three dimensional floating light-field display based on spliced multi-ls," in *SID Symposium Digest of Technical Papers*, vol. 43, pp. 796–799, Wiley Online Library, 2012. (Cited on page 5.)

[14] T.-Y. Chung, J.-Y. Sim, and C.-S. Kim, "Bit allocation algorithm with novel view synthesis distortion model for multiview video plus depth coding," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3254–3267, 2014. (Cited on page 5.)

[15] H.-S. Kim, K.-M. Jeong, S.-I. Hong, N.-Y. Jo, and J.-H. Park, "Analysis of image distortion based on light ray field by multi-view and horizontal parallax only integral imaging display," *Optics express*, vol. 20, no. 21, pp. 23755–23768, 2012. (Cited on page 5.)

[16] A. Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L. D. Soares, "Immersive 3d holoscopic video system," *IEEE Multimedia*, vol. 20, no. 1, pp. 28–37, 2013. (Cited on page 5.)

[17] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: sensing, display, and applications [invited]," *Applied Optics*, vol. 52, no. 4, pp. 546–560, 2013.

[18] A. Stern and B. Javidi, "Three-dimensional image sensing, visualization, and processing using integral imaging," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 591–607, 2006. (Cited on page 5.)

[19] F. P. M. R. D. H. J. PÃ©rez, E. Magdaleno and J. Corrales, "Super-resolution in plenoptic cameras using fpgas," *Sensors (Basel, Switzerland)*, vol. 14, no. 5, pp. 8669–8685, 2014. (Cited on pages 6 and 7.)

[20] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014. (Cited on page 28.)

[21] V. Popescu, B. Benes, P. Rosen, J. Cui, and L. Wang, "A flexible pinhole camera model for coherent nonuniform sampling," 2014. (Cited on page 6.)

[22] C. Birklbauer and O. Bimber, "Panorama light-field imaging," *Computer Graphics Forum*, vol. 33, no. 2, pp. 43–52, 2014. (Cited on page 6.)

[23] S. M. Ayatollahi, A. M. E. Moghadam, and M. S. Hosseini, "A taxonomy of depth map creation methods used in multiview video compression," *Multimedia Tools and Applications*, vol. 72, no. 2, pp. 1887–1909, 2014. (Cited on pages 7 and 149.)

[24] T. Matsuyama, S. Nobuhara, T. Takai, and T. Tung, *Multi-camera Systems for 3D Video Production*, pp. 17–44. 3D Video and Its Applications, Springer, 2012. (Cited on pages 8, 71, and 80.)

[25] S. Zinger, L. Do, and P. H. N. D. With, "Free-viewpoint depth image based rendering," *Journal of Visual Communication and Image Representation*, vol. 21, pp. 533–541, 07 2010. M3: Article. (Cited on pages 8, 30, and 57.)

[26] T. Moons, L. V. Gool, and M. Vergauwen, *3d Reconstruction from Multiple Images: Part 1: Principles*. Now Publishers Inc, 2009. (Cited on page 8.)

[27] Y. Ming and Z. Hu, "Modeling stereopsis via markov random field," *Neural computation*, vol. 22, no. 8, pp. 2161–2191, 2010. (Cited on page 19.)

[28] C. Wheatstone, "Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838. (Cited on page 19.)

[29] I. P. Howard and B. J. Rogers, *Binocular vision and stereopsis*, vol. no. 29. Oxford: Oxford University Press, 1995. (Cited on page 19.)

[30] L. J. Miller, W. Mittenberg, V. M. Carey, M. A. McMorrow, T. E. Kushner, and J. M. Weinstein, "Astereopsis caused by traumatic brain injury," *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, vol. 14, no. 6, pp. 537–543, 1999. (Cited on page 19.)

[31] G. F. Poggio and T. Poggio, "The analysis of stereopsis," *Annual Review of Neuroscience*, vol. 7, no. 1, pp. 379–412, 1984. (Cited on page 19.)

[32] R. Zone, *Stereoscopic Cinema and the origins of 3-D Film, 1838-1952*. University Press of Kentucky, 2014. (Cited on page 19.)

[33] L. Yang, M. P. Tehrani, T. Fujii, and M. Tanimoto, "High-quality virtual view synthesis in 3dtv and ftv," *3D Research*, vol. 2, no. 4, pp. 1–13, 2011. (Cited on page 20.)

[34] D. Min, D. Kim, S. U. Yun, and K. Sohn, "2d/3d freeview video generation for 3dtv system," *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 31–48, 2009.

[35] I. Daribo and H. Saito, "A novel inpainting-based layered depth video for 3dtv," *IEEE Transactions on Broadcasting*, vol. 57, pp. 533–541, 06/15 2011. M3: Article.

[36] A. Koz, C. Cigla, and A. A. Alatan, "Watermarking of free-view video," *IEEE Transactions on Image Processing*, vol. 19, pp. 1785–1797, 07 2010. M3: Article.

[37] W. Li, J. Zhou, B. Li, and M. I. Sezan, "Virtual view specification and synthesis for free viewpoint television," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 533–546, 04 2009. M3: Article. (Cited on page 30.)

[38] F. Sun, G. Jiang, Y. Zhang, M. Yu, Z. Peng, and F. Shao, "Depth no-synthesis-error time-consistent preprocessing for depth coding in fvv system," in *2012 International Workshop on Information and Electronics Engineering, IWIEE 2012*, vol. 29, pp. 2572–2577, 10 March 2012 through 11 March 2012 2012. (Cited on page 22.)

[39] M. Sharma, S. Chaudhury, and B. Lall, "A free viewpoint 3dtv system based on parameterized variety model," pp. 1026–1029, IEEE, 2012.

[40] X. Cao, Y. Liu, X. Ji, and Q. Dai, "Vision field capture for advanced 3dtv applications," pp. 1–4, IEEE, 2011.

[41] S. B. Kang, C. L. Zitnick, M. Uyttendaele, S. Winder, and R. Szeliski, "Free-viewpoint video with stereo and matting," in *Picture Coding Symposium 2004*, pp. 99–104, 15 December 2004 through 17 December 2004 2004. Sponsors: Microsoft Research, Visiowave; IEEE CAS; IEEE SP; Conference code: 64672. (Cited on page 20.)

[42] B. Mendiburu, *3d TV and 3d cinema: tools and processes for creative stereoscopy*. Taylor and Francis, 2011. (Cited on pages 21, 38, and 71.)

[43] J. Zhong, W. B. Kleijn, and X. Hu, "Camera control in multi-camera systems for video quality enhancement," *IEEE Sensors Journal*, vol. 14, no. 9, pp. 2955–2966, 2014. (Cited on page 21.)

[44] D. Chrysostomou and A. Gasteratos, "Optimum multi-camera arrangement using a bee colony algorithm," in *Imaging Systems and Techniques (IST), 2012 IEEE International Conference on*, pp. 387–392, IEEE, 2012. (Cited on pages 74 and 187.)

[45] J. Evers-Senne and R. Koch, "Image based interactive rendering with view dependent geometry," *Computer Graphics Forum*, vol. 22, pp. 573–582, 09 2003. M3: Article.

[46] J. Evers-Senne and R. Koch, "Image-based rendering of complex scenes from a multi-camera rig," *IEE Proceedings – Vision, Image and Signal Processing*, vol. 152, pp. 470–480, 08 2005. M3: Article. (Cited on page 30.)

[47] C. Y. Lee, S. J. Lin, C. W. Lee, and C. S. Yang, "An efficient continuous tracking system in real-time surveillance application," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 1067–1073, 2012. (Cited on page 21.)

[48] X. Jiao, X. Zhao, Y. Yang, Z. Fang, and X. Yuan, "Elemental images correction of camera array pick-up for three-dimensional integral imaging," *Zhongguo Jiguang/Chinese Journal of Lasers*, vol. 39, no. 3, 2012. (Cited on page 22.)

[49] S.-C. Chan, Z.-F. Gan, K.-T. Ng, K.-L. Ho, and H.-Y. Shum, "An object-based approach to image/video-based synthesis and processing for 3-d and multiview televisions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 821–831, 06 2009. M3: Article.

[50] W.-S. Jang and Y.-S. Ho, "Efficient depth map generation with occlusion handling for various camera arrays," *Signal, Image and Video Processing*, vol. 8, no. 2, pp. 287–297, 2014. (Cited on page 22.)

[51] X. Jiang and M. Lambers, "Synthesis of stereoscopic 3d videos by limited resources of range images," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 2, pp. 1220–1224, 2006. ID: 21. (Cited on page 22.)

[52] X. Jiang and M. Lambers, "Dibr-based 3d videos using non video rate range image stream," in *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 1873–1876, 2006. ID: 79.

[53] H.-Y. Shum, S. B. Kang, and S.-C. Chan, "Survey of image-based representations and compression techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1020–1037, 2003. (Cited on page 41.)

[54] Y. K. Park, K. Jung, Y. Oh, G. Lee, S. Lee, H. Lee, J. Kim, J. K. Kim, K. Yun, and N. Hur, "Depth-image-based rendering for 3dtv service over t-dmb," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 122–136, 2009.

[55] J. Park, J.-Y. Choi, I. Ryu, and J.-I. Park, "Universal view synthesis unit for glassless 3dtv," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 706–711, 2012.

[56] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 600–608, 2004.

[57] C. Zhang and T. Chen, "A survey on image-based rendering-representation, sampling and compression," *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[58] J. Kopf, F. Langguth, D. Scharstein, R. Szeliski, and M. Goesele, "Image-based rendering in the gradient domain," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 1–9, 2013. (Cited on page 22.)

[59] C. Gilliam, P.-L. Dragotti, and M. Brookes, "On the spectrum of the plenoptic function," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 502–516, 2014. (Cited on pages 22 and 42.)

[60] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 39–46, ACM, 1995. (Cited on pages 24, 28, 41, 42, and 53.)

[61] H. Y. Shum and S. B. Kang, "A review of image-based rendering techniques," *IEEE/SPIE Visual Communications and Image Processing (VCIP)*, vol. 213, 2000. (Cited on page 24.)

[62] Y. Mao, G. Cheung, A. Ortega, and Y. Ji, "Expansion hole filling in depth-image-based rendering using graph-based interpolation," pp. 1859–1863, IEEE, 2013. (Cited on page 24.)

[63] H.-A. Hsu, C.-K. Chiang, and S.-H. Lai, "Spatio-temporally consistent view synthesis from video-plus-depth data with global optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 1, pp. 74–84, 2014.

[64] S. Lee, S. Lee, H. Wey, J. Lee, and D. Park, "Depth resampling for mixed resolution multiview 3d videos," pp. 827–832, IEEE, 2013.

[65] M. Sharma, S. chaudhury, and B. Lall, "3dtv view generation with virtual pan/tilt/zoom functionality," pp. 1–8, ACM, 2012.

[66] S. Shi, K. Nahrstedt, and R. Campbell, "A real-time remote rendering system for interactive mobile graphics," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 8, no. 3s, pp. 1–20, 2012.

[67] A. T. Tran and K. Harada, "View synthesis with depth information based on graph cuts for ftv," pp. 289–294, IEEE, 2013.

[68] J. Duan and J. Li, "Compression of the layered depth image," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 12, no. 3, pp. 365–372, 2003.

[69] H. Kim, S. Kim, B. Koo, and B. Choi, "Layered-depth image using pixel grouping," pp. 121–127, 2001.

[70] M. Niebner, H. Schafer, and M. Stamminger, "Fast indirect illumination using layered depth images," *The Visual Computer*, vol. 26, no. 6, pp. 679–686, 2010.

[71] U. Takyar, T. Maugey, and P. Frossard, "Multiview navigation based on extended layered depth image representation," 2013. (Cited on page 24.)

[72] G. Medioni and S. B. Kang, *Emerging topics in computer vision*. Prentice Hall PTR, 2004. (Cited on pages 24 and 64.)

[73] P. Merkle, K. Muller, and T. Wiegand, "3d video: acquisition, coding, and display," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 2, pp. 946–950, 2010. (Cited on pages 24, 71, and 74.)

[74] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 8, pp. 993–1008, 2003. (Cited on pages 25 and 26.)

[75] H. Hirschmuller, P. R. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *International Journal of Computer Vision*, vol. 47, no. 1-3, p. 229, 2002. (Cited on page 26.)

[76] R. K. Gupta and S.-Y. Cho, *A Correlation-Based Approach for Real-Time Stereo Matching*, vol. 6454, pp. 129–138. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[77] S. Lefebvre, S. Ambellouis, and F. Cabestaing, "A 1d approach to correlation-based stereo matching," *Image and Vision Computing*, vol. 29, no. 9, pp. 580–593, 2011.

[78] J. Ding, X. Du, X. Wang, and J. Liu, "Improved real-time correlation-based fpga stereo vision system," in *Mechatronics and Automation (ICMA), 2010 International Conference on*, pp. 104–108, IEEE, 2010.

[79] J.-C. Cheng and S.-J. Feng, "A real-time multiresolutional stereo matching algorithm," vol. 3, pp. III–373–6, 2005.

[80] P. N. Hong and C. W. Ahn, "Stereo matching using fusion of spatial weight variable window and adaptive support weight," *International Journal of Computer and Electrical Engineering*, vol. 6, no. 3, pp. 211–217, 2014.

[81] N. Ortigosa and S. Morillas, "Fuzzy free path detection from disparity maps by using least-squares fitting to a plane," *Journal of Intelligent and Robotic Systems*, vol. 75, no. 2, pp. 313–330, 2014.

[82] G.-T. Michailidis, R. Pajarola, and I. Andreadis, "High performance stereo system for dense 3-d reconstruction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 6, pp. 929–941, 2014.

[83] R. K. Gupta and S.-Y. Cho, "Window-based approach for fast stereo correspondence," *IET Computer Vision*, vol. 7, no. 2, p. 123, 2013.

[84] S. Yoon, S.-K. Park, S. Kang, and Y. K. Kwak, "Fast correlation-based stereo matching with the reduction of systematic errors," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2221–2231, 2005. (Cited on page 26.)

[85] Y. Wang and P. Bhattacharya, "Hierarchical stereo correspondence using features of gray connected components," vol. 3, pp. 264–267 vol.3, 1997. (Cited on page 26.)

[86] V. Venkateswar and R. Chellappa, "Hierarchical stereo and motion correspondence using feature groupings," *International Journal of Computer Vision*, vol. 15, no. 3, pp. 245–269, 1995.

[87] K. E. Price, "Hierarchical matching using relaxation," *Computer Vision, Graphics and Image Processing*, vol. 34, no. 1, pp. 66–75, 1986. (Cited on page 26.)

[88] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 181–200, 1999. (Cited on page 26.)

[89] I. Cox, S. Hingorani, S. Rao, and B. Maggs, "A maximum likelihood stereo algorithm," *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 542–542, 1996.

[90] A. P. Harrison and D. Joseph, "Maximum likelihood estimation of depth maps using photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1368–1380, 2012.

[91] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, “Photometric stereo using sparse bayesian regression for general diffuse surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 9, pp. 1816–1831, 2014.

[92] A. Jagmohan, M. Singh, and N. Ahuja, “Dense stereo matching using kernel maximum likelihood estimation,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, pp. 28–31 Vol.3, 2004.

[93] J. Sun, N.-N. Zheng, and H.-Y. Shum, “Stereo matching using belief propagation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 7, pp. 787–800, 2003.

[94] J. Joglekar, S. S. Gedam, and B. K. Mohan, “Image matching using sift features and relaxation labeling technique-a constraint initializing method for dense stereo matching,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5643–5652, 2014.

[95] G. Panahandeh and M. Jansson, “Vision-aided inertial navigation based on ground plane feature detection,” *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 4, pp. 1206–1215, 2014.

[96] S. H. Lee and S. Sharma, “Real-time disparity estimation algorithm for stereo camera systems,” *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1018–1026, 2011.

[97] C. H. Tong, S. Anderson, H. Dong, and T. D. Barfoot, “Pose interpolation for laser-based visual odometry,” *Journal of Field Robotics*, vol. 31, no. 5, pp. 787–813, 2014.

[98] W. Xiong and B. Funt, “Stereo retinex,” *Image and Vision Computing*, vol. 27, no. 1, pp. 178–188, 2009. (Cited on page 26.)

[99] V. Castaneda, D. Mateus, and N. Navab, "Stereo time-of-flight with constructive interference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1402–1413, 2014. (Cited on page 26.)

[100] P. Aflaki, M. Hannuksela, D. Rusanovskyy, and M. Gabbouj, "Nonlinear depth map resampling for depth-enhanced 3-d video coding," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 87–90, 2013. (Cited on page 27.)

[101] M. Gevrekci and K. Pakin, "Depth map super resolution," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 3449–3452, 2011.

[102] J. Kim, G. Jeon, and J. Jeong, "Joint-adaptive bilateral depth map upsampling," *Signal Processing: Image Communication*, vol. 29, no. 4, p. 506, 2014.

[103] K. Muller, P. Merkle, and T. Wiegand, "3-d video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, 2011.

[104] F. Qi, J. Han, P. Wang, G. Shi, and F. Li, "Structure guided fusion for depth map inpainting," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 70–76, 2013.

[105] M.-Y. Shieh and T.-M. Hsieh, "Fast facial detection by depth map analysis," *Mathematical Problems in Engineering*, vol. 2013, pp. 1–10, 2013.

[106] Y. S. Kang, S. B. Lee, and Y. S. Ho, "Depth map upsampling using depth local features," *Electronics Letters*, vol. 50, no. 3, p. 170, 2014.

[107] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference and learning in computer vision and image understanding: A survey," *Computer Vision and Image Understanding*, vol. 117, no. 11, p. 1610, 2013.

[108] Q. Zhou, J. Zhu, and W. Liu, "Learning dynamic hybrid markov random field for image labeling," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 6, pp. 2219–2232, 2013.

[109] S. Yousefi, N. Kehtarnavaz, and Y. Cao, "Computationally tractable stochastic image modeling based on symmetric markov mesh random

fields," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 6, pp. 2192–2206, 2013.

[110] X. Xu, Y. Guo, and Z. Wang, "Cloud image detection based on markov random field," *Journal of Electronics (China)*, vol. 29, no. 3, pp. 262–270, 2012.

[111] A. Dawoud and A. Netchaev, "Preserving objects in markov random fields region growing image segmentation," *Pattern Analysis and Applications*, vol. 15, no. 2, pp. 155–161, 2012. (Cited on page 27.)

[112] H.-Y. Shum, S.-C. Chan, and S. B. Kang, *Image-based rendering*. Springer, 2008. (Cited on page 27.)

[113] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 1, pp. 3–20, 1991. (Cited on pages 27 and 41.)

[114] P. Hariharan, *Optical Holography: Principles, techniques and applications*. Cambridge University Press, 1996. (Cited on page 28.)

[115] C. Birklbauer, S. Opelt, and O. Bimber, "Rendering gigaray light fields," *Computer Graphics Forum*, vol. 32, no. 2pt4, pp. 469–478, 2013. (Cited on page 28.)

[116] C. Birklbauer and O. Bimber, "Light-field supported fast volume rendering," pp. 1–1, ACM, 2012.

[117] S. C. Chan and H. Y. Shum, "A spectral analysis for light field rendering," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 2, pp. 25–28 vol.2, 2000.

[118] D. Horn and B. Chen, "Lightshop: interactive light field manipulation and rendering," pp. 121–128, ACM, 2007.

[119] H. H. Wang, H. H. Wang, M. M. Sun, M. M. Sun, and R. R. Yang, "Space-time light field rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 4, pp. 697–710, 2007.

[120] J. Lehtinen, T. Aila, J. Chen, S. Laine, and F. Durand, "Temporal light field reconstruction for rendering distribution effects," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 1, 2011.

[121] M. Levoy and P. Hanrahan, "Light field rendering," pp. 31–42, ACM, 1996. (Cited on page 76.)

[122] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, no. 2, pp. 121–138, 2004.

[123] M. Escriva, J. Blasco, F. Abad, E. Camahort, and R. Vivo, "Autostereoscopic rendering of multiple light fields," *Computer Graphics Forum*, vol. 28, no. 8, p. 2057, 2009.

[124] S. B. Oh, S. Kashyap, R. Garg, S. Chandran, and R. Raskar, "Rendering wave effects with augmented light field," *Computer Graphics Forum*, vol. 29, no. 2, pp. 507–516, 2010.

[125] K. Takahashi, A. Kubota, and T. Naemura, "A focus measure for light field rendering," vol. 4, pp. 2475–2478 Vol. 4, 2004.

[126] J. van der Linden, "Multiple light field rendering," pp. 197–ff, ACM, 2003.

[127] K. Takahashi and T. Naemura, "Layered light-field rendering with focus measurement," *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 519–530, 2006.

[128] M. Fuchs, M. Kachele, and S. Rusinkiewicz, "Design and fabrication of faceted mirror arrays for light field capture," *Computer Graphics Forum*, vol. 32, no. 8, pp. 246–257, 2013. (Cited on page 28.)

[129] S. M. Seitz and C. R. Dyer, "Physically-valid view synthesis by image interpolation," in *Representation of Visual Scenes, 1995.(In Conjuction with ICCV'95), Proceedings IEEE Workshop on*, pp. 18–25, IEEE, 1995. (Cited on page 28.)

[130] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia*, vol. 4, no. 1, pp. 34–47, 1997. (Cited on page 28.)

[131] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," pp. 425–432, ACM, 2001. (Cited on page 28.)

[132] Y. MATSUSHITA, S. LIN, H.-Y. SHUM, X. TONG, and S. B. KANG, "Lighting and shadow interpolation using intrinsic lumigraphs," *International Journal of Image and Graphics*, vol. 4, no. 4, pp. 585–604, 2004.

[133] Y. Morvan and C. O'sullivan, "A perceptual approach to trimming and tuning unstructured lumigraphs," *ACM Transactions on Applied Perception (TAP)*, vol. 5, no. 4, pp. 1–24, 2009.

[134] Y. Morvan and C. O'Sullivan, "A perceptual approach to trimming unstructured lumigraphs," pp. 61–68, ACM, 2007.

[135] H. Schirmacher, W. Heidrich, and H.-P. Seidel, "Adaptive acquisition of lumigraphs from synthetic scenes," *Computer Graphics Forum*, vol. 18, no. 3, pp. 151–160, 1999.

[136] H. Schirmacher, L. Ming, and H.-P. Seidel, "On-the-fly processing of generalized lumigraphs," *Computer Graphics Forum*, vol. 20, no. 3, pp. 165–174, 2001. (Cited on page 28.)

[137] C. Zhang and J. Li, "On the compression and streaming of concentric mosaic data for free wandering in a realistic environment over the internet," *IEEE Transactions on Multimedia*, vol. 7, no. 6, pp. 1170–1182, 2005. (Cited on page 28.)

[138] Z. Fu and L. Wang, "Optimized design of automatic image mosaic," *Multimedia Tools and Applications*, vol. 72, no. 1, pp. 503–514, 2014. (Cited on page 28.)

[139] D. S. Katz, G. B. Berriman, and R. G. Mann, "Collaborative astronomical image mosaics," 2010. (Cited on page 28.)

[140] P. Chatkaewmanee and M. N. Dailey, "Object virtual viewing using adaptive tri-view morphing," *IET Image Processing*, vol. 7, no. 6, p. 586, 2013. (Cited on page 28.)

[141] S.-M. Rhee, J. Choi, and U. Neumann, *Stereoscopic View Synthesis by View Morphing*, vol. 5359, pp. 924–933. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[142] S. Seitz and C. Dyer, "View morphing," pp. 21–30, ACM, 1996.

[143] W. Yang and J. Feng, "2d shape morphing via automatic feature matching and hierarchical interpolation," *Computers and Graphics*, vol. 33, no. 3, pp. 414–423, 2009.

[144] J. Y. Kang and B. Lee, "Mesh-based morphing method for rapid hull form generation," *Computer-Aided Design*, vol. 42, no. 11, pp. 970–976, 2010. (Cited on page 28.)

[145] S. Avidan and A. Shashua, "Novel view synthesis by cascading trilinear tensors," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, no. 4, pp. 293–306, 1998. (Cited on page 29.)

[146] S. Avidan and A. Shashua, "Novel view synthesis in tensor space," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 1034–1040, 1997.

[147] H. Li and R. Hartley, "Inverse tensor transfer for novel view synthesis," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, pp. II–97–100, 2005.

[148] Q. Yu, Q. Liao, and Z. Lu, "Six-point camera pose estimation algorithm applied into tensor-based novel view synthesis," in *Imaging Systems and Techniques, 2009. IST'09. IEEE International Workshop on*, pp. 205–209, 2009.

[149] H. Li and R. Hartley, "Inverse tensor transfer with applications to novel view synthesis and multi-baseline stereo," *Signal Processing: Image Communication*, vol. 21, no. 9, pp. 724–738, 2006.

[150] M. A. O. Vasilescu, M. A. O. Vasilescu, and D. Terzopoulos, "Multilinear (tensor) image synthesis, analysis, and recognition [exploratory dsp]," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 118–123, 2007. (Cited on page 29.)

[151] M. Goesele, J. Ackermann, S. Fuhrmann, C. Haubold, R. Klowsky, and T. Darmstadt, "Ambient point clouds for view interpolation," *ACM Transactions on Graphics*, vol. 29, no. 4, p. 1, 2010. (Cited on page 29.)

[152] M. Makar, Y.-C. Lin, N.-M. Cheung, D. Pang, and B. Girod, "Quality-controlled view interpolation for multiview video," pp. 1805–1808, IEEE, 2011.

[153] P. K. Rana, P. K. Rana, M. Flierl, and M. Flierl, "Depth consistency testing for improved view interpolation," pp. 384–389, IEEE, 2010.

[154] T. Stich, C. Linz, G. Albuquerque, and M. Magnor, "View and time interpolation in image space," *Computer Graphics Forum*, vol. 27, no. 7, pp. 1781–1787, 2008.

[155] K. Takahashi, "Theoretical analysis of view interpolation with inaccurate depth information," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 718–732, 2012.

[156] X. Xiu, D. Pang, and J. Liang, "Rectification-based view interpolation and extrapolation for multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 693–707, 2011.

[157] X. Sun and E. Dubois, "A matching-based view interpolation scheme," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 2, pp. ii–877, 2005.

[158] S. Ince and J. Konrad, "Occlusion-aware view interpolation," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–15, 2008 2008. M3: Article.

[159] H. Bao, L. Chen, J. Ying, and Q. Peng, “Non-linear view interpolation,” *The Journal of Visualization and Computer Animation*, vol. 10, no. 4, pp. 233–241, 1999.

[160] G. P. Fickel, C. R. Jung, T. Malzbender, R. Samadani, and B. Culbertson, “Stereo matching and view interpolation based on image domain triangulation,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3353–3365, 2013. (Cited on page 29.)

[161] L. Zhang and W. J. Tam, “Stereoscopic image generation based on depth images for 3d tv,” *Broadcasting, IEEE Transactions on*, vol. 51, no. 2, pp. 191–199, 2005. (Cited on pages 30, 31, and 56.)

[162] N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “Free viewpoint image generation using multi-pass dynamic programming,” in *Electronic Imaging 2007*, pp. 64901F–64901F–11, International Society for Optics and Photonics, 2007. (Cited on page 30.)

[163] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Multi-view video plus depth representation and coding,” in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I–201–I–204, IEEE, 2007. (Cited on page 30.)

[164] A. Fusiello, “Specifying virtual cameras in uncalibrated view synthesis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 604–611, 05 2007. M3: Article. (Cited on page 30.)

[165] S. Chan, H. Y. Shum, and K. T. Ng, “Image-based rendering and synthesis,” *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 22–33, 2007. (Cited on pages 30, 34, 41, 42, 53, and 80.)

[166] A. Fusiello and L. Irsara, “An uncalibrated view-synthesis pipeline,” in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pp. 609–614, IEEE, 2007. (Cited on pages 30 and 148.)

[167] E. Lee, Y. Kang, Y. Jung, and Y. Ho, "3-d video generation using hybrid camera system," in *Proceedings of the 2nd International Conference on Immersive Telecommunications*, p. 5, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009. (Cited on page 31.)

[168] M. Farre, O. Wang, M. Lang, N. Stefanoski, A. Hornung, and A. Smolic, "Automatic content creation for multiview autostereoscopic displays using image domain warping," pp. 1–6, IEEE, 2011. (Cited on page 31.)

[169] A. Speers and M. Jenkin, "Tuning stereo image matching with stereo video sequence processing," in *Joint International Conference on Human-Centered Computer Environments, HCCE 2012*, pp. 208–214, 8 March 2012 through 13 March 2012 2012. (Cited on page 32.)

[170] D. N. Bhat and S. K. Nayar, "Ordinal measures for image correspondence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 4, pp. 415–423, 1998. (Cited on page 32.)

[171] J. Lewis, "Fast normalized cross-correlation," in *Vision interface*, vol. 10, pp. 120–123, 1995. (Cited on page 33.)

[172] F. A. Kingdom, "Binocular vision: The eyes add and subtract," *Current Biology*, vol. 22, no. 1, pp. R22–R24, 2012. (Cited on page 34.)

[173] S. Hadjitheophanous, C. Ttofis, A. Georghiades, and T. Theocharides, "Towards hardware stereoscopic 3d reconstruction a real-time fpga computation of the disparity map," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2010*, pp. 1743–1748, IEEE, 2010. (Cited on page 34.)

[174] S. Bae, H. Lee, H. Park, H. Cho, J. Park, and J. Kim, "The effects of egocentric and allocentric representations on presence and perceived realism: Tested in stereoscopic 3d games," *Interacting with Computers*, 2012. (Cited on page 34.)

[175] C. Lee, Y.-S. Ho, and B. Choi, “Efficient multiview depth video coding using depth synthesis prediction,” *Optical Engineering*, vol. 50, pp. 077004–077004, 07 2011. M3: Article. (Cited on pages 34 and 41.)

[176] C. G. Gurler, K. T. Bagci, and A. M. Tekalp, “Adaptive stereoscopic 3d video streaming,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 2409–2412, IEEE, 2010. (Cited on page 34.)

[177] G. B. Akar, G. B. Akar, A. M. Tekalp, C. Fehn, and M. R. Civanlar, “Transport methods in 3dtv-a survey,” vol. 17, pp. 1622–1630, IEEE, 2007.

[178] G. V. Wallendael, S. V. Leuven, J. D. Cock, F. Bruls, and de Walle Van, “3d video compression based on high efficiency video coding,” *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 137–145, 02 2012. M3: Article.

[179] A. Vetro, A. M. Tourapis, K. Muller, and T. Chen, “3d-tv content storage and transmission,” *IEEE Transactions on Broadcasting*, vol. 57, pp. 384–394, 06/15 2011. M3: Article. (Cited on page 34.)

[180] S. K. Nayar, “Catadioptric omnidirectional camera,” in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 482–488, IEEE, 1997. (Cited on pages 35 and 54.)

[181] G. R. Jones, D. Lee, N. S. Holliman, and D. Ezra, “Controlling perceived depth in stereoscopic images,” in *Photonics West 2001-Electronic Imaging*, pp. 42–53, International Society for Optics and Photonics, 2001. (Cited on pages 35, 37, 38, and 187.)

[182] L. E. Gurrieri and E. Dubois, “Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence,” in *IST/SPIE Electronic Imaging*, pp. 82882D–82882D–14, International Society for Optics and Photonics, 2012. (Cited on pages 35 and 44.)

[183] R. Szeliski and H. Y. Shum, “Creating full view panoramic image mosaics and environment maps,” in *Proceedings of the 24th annual conference on Com-*

puter graphics and interactive techniques, pp. 251–258, ACM Press/Addison-Wesley Publishing Co., 1997. (Cited on page 45.)

[184] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006. (Cited on pages 35, 43, 45, and 49.)

[185] R. F. van der Willigen, W. M. Harmening, S. Vossen, and H. Wagner, "Disparity sensitivity in man and owl: Psychophysical evidence for equivalent perception of shape-from-stereo," *Journal of Vision*, vol. 10, no. 1, 2010. (Cited on page 36.)

[186] K. Iizuka, *Engineering optics*, vol. 35. Springer, 2008. (Cited on pages 36, 37, 83, 84, and 211.)

[187] L. Leroy, P. Fuchs, and G. Moreau, "Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 10, pp. 3998–4004, 2012. (Cited on page 37.)

[188] G. Leon, H. Kalva, and B. Furht, "3d video quality evaluation with depth quality variations," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pp. 301–304, 2008. ID: 58. (Cited on page 37.)

[189] M. Wopking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *Journal of the Society for Information Display*, vol. 3, no. 3, pp. 101–103, 1995. (Cited on page 38.)

[190] P. Corke, *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*, vol. 73. Springer, 2011. (Cited on pages 40, 47, 110, and 112.)

[191] A. Lumsdaine, G. Chunev, and T. Georgiev, "Plenoptic rendering with interactive performance using gpus," in *Image Processing: Algorithms and*

Systems X; and Parallel Processing for Imaging Applications II, vol. 8295, 23 January 2012 through 25 January 2012 2012. (Cited on page 41.)

[192] C. Zhang and T. C. 1965, *Light field sampling electronic resource*, vol. 6. San Rafael, Calif. (Fourth Street, San Rafael, CA1 USA): Morgan and Claypool Publishers, 1st ed., c2006.

[193] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational plenoptic imaging," *Computer Graphics Forum*, vol. 30, pp. 2397–2426, 12 2011. M3: Article. (Cited on page 41.)

[194] C. Zhang, E. Dubois, and Y. Zhao, "Virtual cubic panorama synthesis based on triangular reprojection," *Computer Animation and Virtual Worlds*, vol. 25, no. 2, pp. 143–154, 2014. (Cited on page 42.)

[195] P. Supan, I. Stuppacher, and M. Haller, "Image based shadowing in real-time augmented reality.," *IJVR*, vol. 5, no. 3, pp. 1–7, 2006. (Cited on page 42.)

[196] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 73–82, ACM Press/Addison-Wesley Publishing Co., 1999. (Cited on page 42.)

[197] A. Bartoli, N. Dalal, and R. Horaud, "Motion panoramas," *Computer Animation and Virtual Worlds*, vol. 15, no. 5, pp. 501–517, 2004. (Cited on page 42.)

[198] M. GONG and Y.-H. YANG, "Rayset: A taxonomy for image-based rendering," *International Journal of Image and Graphics*, vol. 6, no. 3, pp. 313–339, 2006. (Cited on pages 44 and 78.)

[199] Y. Cui, N. Hasler, T. Thormahlen, and H. P. Seidel, "Scale invariant feature transform with irregular orientation histogram binning," *Image Analysis and Recognition*, pp. 258–267, 2009. (Cited on pages 45, 123, and 125.)

[200] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II–506–II–513 Vol. 2, IEEE, 2004. (Cited on page 45.)

[201] C. J. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 3, pp. 206–218, 1997. (Cited on page 46.)

[202] S. Baker and S. K. Nayar, "A theory of catadioptric image formation," in *Computer Vision, 1998. Sixth International Conference on*, pp. 35–42, IEEE, 1998. (Cited on page 47.)

[203] C.-K. Sung and C.-H. Lu, "Single-camera panoramic stereo system model based on skew ray tracing," *Optik - International Journal for Light and Electron Optics*, vol. 123, pp. 594–603, 04 2012. M3: Article. (Cited on pages 47, 49, and 54.)

[204] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," *Computer Vision-ECCV 2004*, pp. 377–389, 2004. (Cited on pages 48, 55, and 126.)

[205] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007. (Cited on pages 48 and 78.)

[206] E. W. Weisstein, "Cylindrical projection." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CylindricalProjection.html. (Cited on page 49.)

[207] J. P. Siebert and S. J. Marshall, "Human body 3d imaging by speckle texture projection photogrammetry," *Sensor Review*, vol. 20, no. 3, pp. 218–226, 2000. (Cited on page 52.)

[208] H.-G. Maas, T. Putze, and P. Westfeld, *Recent developments in 3D-PTV and Tomo-PIV*, pp. 53–62. Imaging Measurement Methods for Flow Analysis, Springer, 2009.

[209] P. Jackman, D. W. Sun, and G. ElMasry, "Robust colour calibration of an imaging system using a colour space transform and advanced regression modelling," *Meat Science*, vol. 91, no. 4, pp. 402–407, 2012. (Cited on page 52.)

[210] X. Li, B. Liu, and E. Wu, "Full solid angle panoramic viewing by depth image warping on field programmable gate array," *The International Journal of Virtual Reality*, vol. 6, no. 2, pp. 69–77, 2011. (Cited on page 53.)

[211] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams," *Computer Vision and Image Understanding*, vol. 71, no. 2, pp. 154–165, 1998. (Cited on page 53.)

[212] F. Isgro, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 3, pp. 288–303, 2004. (Cited on page 53.)

[213] B. MicusiÂk and T. Pajdla, *Omnidirectional Camera Model and Epipolar Geometry Estimation by RANSAC with Bucketing?*, pp. 83–90. Image Analysis, Springer, 2003. (Cited on page 54.)

[214] F. Wyrowski and M. Kuhn, "Introduction to field tracing," *Journal of Modern Optics*, vol. 58, pp. 449–466, 03/10 2011. M3: Article. (Cited on page 55.)

[215] Y. Xiong and K. Pulli, "Sequential image stitching for mobile panoramas," in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, pp. 1–5, IEEE, 2009. (Cited on page 56.)

[216] A. I. Audu and A. H. Sadka, "Metric aspect of depth image-based rendering," in *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*, pp. 1–6, IEEE, 2013. (Cited on page 56.)

[217] E. K. Lee and Y. S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-d video," *Journal of Visual Communication and Image Representation*, vol. 22, no. 1, pp. 73–84, 2011. (Cited on pages 57 and 77.)

[218] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217–234, 2007. (Cited on page 57.)

[219] M. I. Lourakis and A. A. Argyros, "Sba: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software (TOMS)*, vol. 36, no. 1, p. 2, 2009. (Cited on pages 58, 59, 63, 65, 67, 68, 71, 170, and 171.)

[220] A. V. Bhavsar and A. N. Rajagopalan, "Towards unrestrained depth inference with coherent occlusion filling," *International Journal of Computer Vision*, vol. 97, no. 2, pp. 167–190, 2012. (Cited on pages 58, 59, and 74.)

[221] N. Borlin and P. Grussenmeyer, "Bundle adjustment with and without damping," *The Photogrammetric Record*, vol. 28, no. 144, pp. 396–415, 2013. (Cited on page 59.)

[222] G. Wang, J. Wu, and Z. Ji, "Single view based pose estimation from circle or parallel lines," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 977–985, 2008. (Cited on page 60.)

[223] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 578–589, 2003.

[224] F. Duan, F. Wu, and Z. Hu, "Pose determination and plane measurement using a trapezium," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 223–231, 2008.

[225] K. Rahbar and H. R. Pourreza, "Inside looking out camera pose estimation for virtual studio," *Graphical Models*, vol. 70, no. 4, pp. 57–75, 2008.

[226] L. Quan and Z. Lan, "Linear n-point camera pose determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 774–780, 1999.

[227] H. F. Ugurdag, S. Goren, and F. Canbay, "Gravitational pose estimation," *Computers and Electrical Engineering*, vol. 36, no. 6, pp. 1165–1180, 2010.

[228] S. Zhang, Y. Ding, K. Hao, and D. Zhang, "An efficient two-step solution for vision-based pose determination of a parallel manipulator," *Robotics and Computer Integrated Manufacturing*, vol. 28, no. 2, p. 182, 2012. (Cited on page 60.)

[229] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, *Bundle adjustment - a modern synthesis*, pp. 298–372. Vision algorithms: theory and practice, Springer, 2000. (Cited on pages 60, 61, 64, and 65.)

[230] M. yi Shen, Z. yu Xiang, and J. lin Liu, "Vision based terrain reconstruction for planet rover using a special binocular bundle adjustment," *Journal of Zhejiang University SCIENCE A*, vol. 9, no. 10, pp. 1341–1350, 2008.

[231] L. Reyes and E. Bayro-Corrochano, "Simultaneous and sequential reconstruction of visual primitives with bundle adjustment," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 63–78, 2006.

[232] M. Lhuillier, "Incremental fusion of structure-from-motion and gps using constrained bundle adjustments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2489–2495, 2012.

[233] R. I. Hartley, *Euclidean reconstruction from uncalibrated views*, pp. 235–256. Applications of invariance in computer vision, Springer, 1994.

[234] D. D. Lichti, C. Kim, and S. Jamtsho, "An integrated bundle adjustment approach to range camera geometric self-calibration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 4, pp. 360–368, 2010.

[235] B. Fodor, C. Kazo, Z. Janko, and L. Hajder, "Normal map recovery using bundle adjustment," *IET Computer Vision*, vol. 8, no. 1, p. 66, 2014.

[236] K. F. Blonquist and R. T. Pack, "A bundle adjustment approach with inner constraints for the scaled orthographic projection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 6, p. 919, 2011. (Cited on pages 60 and 63.)

[237] X. Liu, W. Gao, and Z.-Y. Hu, "Hybrid parallel bundle adjustment for 3d scene reconstruction with massive points," *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1269–1280, 2012. (Cited on pages 60 and 61.)

[238] M. Hu, G. Penney, M. Figl, P. Edwards, F. Bello, R. Casula, D. Rueckert, and D. Hawkes, "Reconstruction of a 3d surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes," *Medical image analysis*, vol. 16, no. 3, pp. 597–611, 2012. (Cited on page 61.)

[239] R. Li, J. Hwangbo, Y. Chen, and K. Di, "Rigorous photogrammetric processing of hirise stereo imagery for mars topographic mapping," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 7, pp. 2558–2572, 2011. (Cited on page 62.)

[240] P. F. McLauchlan and A. Jaenicke, "Image mosaicing using sequential bundle adjustment," *Image and Vision Computing*, vol. 20, no. 9, pp. 751–759, 2002. (Cited on page 62.)

[241] D. Nister, *Frame decimation for structure and motion*, pp. 17–34. 3D Structure from Images-SMILE 2000, Springer, 2001. (Cited on page 62.)

[242] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 8, pp. 1362–1376, 2010. (Cited on page 62.)

[243] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1066–1077, 2008. (Cited on page 62.)

[244] C. Engels, H. Stewenius, and D. Nister, "Bundle adjustment rules," *Photogrammetric computer vision*, vol. 2, 2006. (Cited on page 62.)

[245] D. Sibley, C. Mei, I. Reid, and P. Newman, "Adaptive relative bundle adjustment.," in *Robotics: science and systems*, 2009. (Cited on pages 63 and 66.)

[246] A. Bartoli and P. Sturm, "Structure-from-motion using lines: Representation, triangulation, and bundle adjustment," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 416–441, 2005. (Cited on page 63.)

[247] K. Konilige, "Sparse sparse bundle adjustment," 2010. (Cited on page 63.)

[248] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963. (Cited on page 66.)

[249] G. H. Golub and C. F. van Van Loan, "Matrix computations (johns hopkins studies in mathematical sciences)," 1996. (Cited on page 68.)

[250] R. Cesar, "A pragmatic introduction to machine vision, by r. jain, r. kasturi and b. g. schunck," *Real-Time Imaging*, vol. 1, no. 6, pp. 437–439, 1995. (Cited on page 71.)

[251] G. Lippman, "Epreuves reversibles photographies integrales," *CR Acad.Sci*, vol. 146, pp. 446–451, 1908. (Cited on page 71.)

[252] S. Jarusirisawad and H. Saito, "3dtv view generation using uncalibrated pure rotating and zooming cameras," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 17–30, 2009. (Cited on page 71.)

[253] C. Zhang, "Multiview imaging and 3dtv," *IEEE Signal Processing Magazine*, vol. 1053, no. 5888/07, 2007. (Cited on page 73.)

[254] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 765–776, 2005. (Cited on page 73.)

[255] Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z.-H. Zhou, "Multi-view video summarization," *Multimedia, IEEE Transactions on*, vol. 12, no. 7, pp. 717–729, 2010. (Cited on page 73.)

[256] H. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*. Academic Press, 2009. (Cited on page 73.)

[257] A. Saxena, J. Schulte, and A. Y. Ng, "Depth estimation using monocular and stereo cues.," in *IJCAI*, pp. 2197–2203, 2007. (Cited on page 73.)

[258] Q. Huynh-Thu and L. Schiatti, "Examination of 3d visual attention in stereoscopic video content," in *IST/SPIE Electronic Imaging*, pp. 78650J–78650J–15, International Society for Optics and Photonics, 2011. (Cited on pages 74 and 184.)

[259] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3d," *ACM Transactions on Graphics*, vol. 29, no. 4, p. 1, 2010. (Cited on page 74.)

[260] H.-C. Li, J. Seo, K. Kham, and S. Lee, "Measurement of 3d visual fatigue using event-related potential (erp): 3d oddball paradigm," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*, pp. 213–216, IEEE, 2008. (Cited on page 74.)

[261] R. Schneiders, "Algorithms for quadrilateral and hexahedral mesh generation," *Proceedings of the VKI lecture series on computational fluid cynamics*, 2000. (Cited on page 74.)

[262] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," in *Proceedings of the 13th Eurographics workshop on Rendering*, pp. 77–86, Eurographics Association, 2002. (Cited on page 75.)

[263] C. Zhang and T. Chen, "A self-reconfigurable camera array," in *ACM SIGGRAPH 2004 Sketches*, p. 151, ACM, 2004. (Cited on page 75.)

[264] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," in *IST/SPIE Electronic Imaging*, pp. 78730X–78730X–13, International Society for Optics and Photonics, 2011. (Cited on page 75.)

[265] K. Venkataraman, D. Lelescu, J. Duparre, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: an ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 1–13, 2013. (Cited on page 76.)

[266] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by bound optics (tombo): concept and experimental verification," *Applied Optics*, vol. 40, no. 11, pp. 1806–1813, 2001. (Cited on page 76.)

[267] R. Horisaki, K. Kagawa, Y. Nakao, T. Toyoda, Y. Masaki, and J. Tanida, "Irregular lens arrangement design to improve imaging performance of compound-eye imaging systems," *Applied physics express*, vol. 3, no. 2, p. 022501, 2010. (Cited on page 76.)

[268] W.-S. Jang and Y.-S. Ho, "Direct depth value extraction method for various stereo camera arrangements," pp. 128–131, 2013. (Cited on page 77.)

[269] Y.-S. Ho, "Challenging technical issues of 3d video processing," *Journal of Convergence*, vol. 4, no. 1, pp. 1–6, 2013. (Cited on page 77.)

[270] T. Naemura, J. Tago, and H. Harashima, "Real-time video-based modeling and rendering of 3d scenes," *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 66–73, 2002. (Cited on page 77.)

[271] Y. Taguchi, T. Koike, K. Takahashi, and T. Naemura, "Transcaip: A live 3d tv system using a camera array and an integral photography display with interactive control of viewing parameters," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 5, pp. 841–852, 2009. (Cited on page 78.)

[272] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, 2006. (Cited on page 78.)

[273] M. Tanimoto, "Overview of free viewpoint television," *Signal Processing: Image Communication*, vol. 21, pp. 454–461, 07 2006. M3: Article. (Cited on page 78.)

[274] T. Koyama, I. Kitahara, and Y. Ohta, "Live mixed-reality 3d video in soccer stadium," in *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, pp. 178–186, IEEE, 2003. (Cited on page 78.)

[275] C. Theobalt, J. Carranza, M. Magnor, and H.-P. Seidel, "3d video: being part of the movie," *ACM SIGGRAPH Computer Graphics*, vol. 38, no. 3, pp. 18–20, 2004. (Cited on page 79.)

[276] "Digital imaging," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, no. S1, pp. 3–10, 2008. (Cited on page 84.)

[277] A. Foi, A. Foi, S. Alenius, S. Alenius, V. Katkovnik, V. Katkovnik, K. Egiazarian, and K. Egiazarian, "Noise measurement for raw-data of digital imaging sensors by automatic segmentation of nonuniform targets," *IEEE Sensors Journal*, vol. 7, no. 10, pp. 1456–1461, 2007. (Cited on page 84.)

[278] R. D. Fiete, "Image chain analysis for space imaging systems," *Journal of Imaging Science and Technology*, vol. 51, no. 2, pp. 103–109, 2007. (Cited on page 84.)

[279] R. D. Fiete, *Modeling the imaging chain of digital cameras*. SPIE press, 2010. (Cited on pages 84 and 85.)

[280] K. Chang, *RF and microwave wireless systems*, vol. 161. John Wiley and Sons, 2004. (Cited on page 85.)

[281] H. J. Visser, *Antenna theory and applications*. Hoboken, N.J: Wiley-Blackwell, 2012.

[282] R. S. Elliott, *Antenna theory and design*. Hoboken, N.J: Wiley-Interscience, 2003.

[283] J. M. Blackledge, *Digital Image Processing: Mathematical and Computational*. Horwood publishing, 2005. (Cited on page 85.)

[284] S. J. Orfanidis, *Electromagnetic waves and antennas*. Rutgers University New Brunswick, NJ, 2002. (Cited on pages 87 and 89.)

[285] V. Antti and A. Lehto, *Radio engineering for wireless communication and sensor applications*. Artech House, 2003. (Cited on pages 93 and 96.)

[286] G. Cristobal, P. Schelkens, and H. Thienpont, *Optical and Digital Image Processing: Fundamentals and Applications*. John Wiley and Sons, 2011.

[287] D. J. Daniels, *Antennas*, pp. 83–127. Hoboken, NJ, USA: John Wiley and Sons, Inc. (Cited on page 93.)

[288] C. A. Balanis, *Antenna theory: analysis and design*. John Wiley and Sons, 2012. (Cited on pages 95 and 97.)

[289] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O.Firschein, eds*, pp. 61–62, 1987. (Cited on page 113.)

[290] K. Mikolajczyk and C. Schmid, “Indexing based on scale invariant interest points,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, pp. 525–531, IEEE, 2001. (Cited on page 113.)

[291] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008. (Cited on page 113.)

[292] A. Baumberg, “Reliable feature matching across widely separated views,” in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1, pp. 774–781, IEEE, 2000. (Cited on page 113.)

[293] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981. (Cited on pages 114 and 126.)

[294] 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. (Cited on page 114.)

[295] Q. Chen, K. Kotani, F. Lee, and T. Ohmi, “Robust local image descriptor based on vector quantization histogram,” *International Journal of Digital Content Technology and its Applications*, vol. 6, no. 7, pp. 253–260, 2012. (Cited on page 123.)

[296] Y. J. Zhong and Q. Cai, *A novel registration approach for mammograms based on SIFT and graph transformation*, vol. 157-158 of *Mechatronics and Applied Mechanics*. hong kong ed., 27 December 2011 through 28 December 2011 2012.

[297] C. Wang, H. Yang, Y. Chen, L. Sun, H. Wang, and Y. Zhou, “Identification of image-spam based on perimetric complexity analysis and sift image matching algorithm,” *Journal of Information and Computational Science*, vol. 9, no. 4, pp. 1073–1081, 2012.

[298] Y. K. Han, Y. G. Byun, J. W. Choi, D. Y. Han, and Y. I. Kim, "Automatic registration of high-resolution images using local properties of features," *Photogrammetric Engineering and Remote Sensing*, vol. 78, no. 3, pp. 211–221, 2012.

[299] F. C. Huang, S. Y. Huang, J. W. Ker, and Y. C. Chen, "High-performance sift hardware accelerator for real-time image feature extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 340–351, 2012.

[300] S. Lee, K. T. Mai, and W. Jeong, "Virtual high dynamic range imaging for robust recognition," in *6th International Conference on Ubiquitous Information Management and Communication, ICUIMC'12*, Affiliation: School of Information and Communication Engineering, Department of Interaction Science, Sungkyunkwan University, Suwon, South Korea; Affiliation: Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea; Correspondence Address: Lee, S.; School of Information and Communication Engineering, Department of Interaction Science, Sungkyunkwan University, Suwon, South Korea; email: lsh@ece.skku.ac.kr, 20 February 2012 through 22 February 2012 2012. (Cited on page 123.)

[301] J. G. Fryer and D. C. Brown, "Lens distortion for close-range photogrammetry," *Photogrammetric Engineering and Remote Sensing*, vol. 52, no. 1, pp. 51–58, 1986. (Cited on page 124.)

[302] S. Aritan, "Efficiency of non-linear lens distortion models in biomechanical analysis of human movement," *Measurement (02632241)*, vol. 43, pp. 739–746, 07 2010. M3: Article.

[303] M. Goljan and J. Fridrich, "Sensor-fingerprint based identification of images corrected for lens distortion," in *Media Watermarking, Security, and Forensics 2012*, vol. 8303, 23 January 2012 through 25 January 2012 2012. (Cited on page 124.)

[304] A. Baumberg, "Blending images for texturing 3d models," in *Proceedings of the British Machine Vision Conference*, pp. 404–413, 2002. (Cited on page 126.)

[305] V. Douskos, I. Kalisperakis, and G. Karras, "Automatic calibration of digital cameras using planar chess-board patterns," in *Proceedings of the 8th Conference on Optical*, pp. 9–12, 2007. (Cited on page 127.)

[306] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004. (Cited on page 127.)

[307] E. Cranger and K. Cupery, "An optical merit function (sqf), which correlates with subjective image judgements," *Phot.Sci.Eng*, vol. 16, pp. 221–230, 1972. (Cited on page 138.)

[308] D. L. Lee and A. T. Winslow, "Performance of three image-quality metrics in ink-jet printing of plain papers," *Journal of Electronic Imaging*, vol. 2, pp. 174–184, july 1993. (Cited on page 138.)

[309] S. Roy, J. Meunier, and I. J. Cox, "Cylindrical rectification to minimize epipolar distortion," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 393–399, IEEE, 1997. (Cited on page 145.)

[310] P. K. Jain and C. Jawahar, "Homography estimation from planar contours," in *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pp. 877–884, IEEE, 2006. (Cited on page 146.)

[311] C. Wohler, *3D computer vision*. Springer, 2009. (Cited on page 147.)

[312] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003. (Cited on page 147.)

[313] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–195, 1998. (Cited on page 148.)

[314] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 8, no. 1, pp. 3–30, 1998. (Cited on page 156.)

[315] J.Xiao, "Princeton vision toolkit." Available from: http://vision.princeton.edu/code.html, 2013. (Cited on page 166.)

[316] W.Project, "Ply (file format)." Available from: http://en.wikipedia.org/wiki/PLY-(file-format), 2014. (Cited on page 166.)

[317] M. Lourakis and A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm," *Institute of Computer Science-FORTH*, vol. 1, no. 2, p. 5, 2004. (Cited on page 170.)

[318] M. Kandefer and S. C. Shapiro, "A categorization of contextual constraints.," in *AAAI Fall Symposium: Biologically Inspired Cognitive Architectures*, pp. 88–93, 2008. (Cited on page 170.)

[319] S. Z. Li and S. Singh, *Markov random field modeling in image analysis*, vol. 26. Springer, 2009. (Cited on page 170.)

[320] G. R. Cross and A. K. Jain, "Markov random field texture models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 1, pp. 25–39, 1983. (Cited on page 170.)

[321] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in *Advances in Neural Information Processing Systems*, pp. 1161–1168, 2005. (Cited on page 170.)

[322] A. Saxena, M. Sun, and A. Y. Ng, "3-d reconstruction from sparse views using monocular vision," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007. (Cited on page 171.)

[323] A. Saxena, S. H. Chung, and A. Y. Ng, "3-d depth reconstruction from a single still image," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 53–69, 2008.

[324] A. Saxena, M. Sun, and A. Y. Ng, "Learning 3-d scene structure from a single still image," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007. (Cited on page 170.)

[325] G. Zhang, X. Qin, W. Hua, T.-T. Wong, P.-A. Heng, and H. Bao, "Robust metric reconstruction from challenging video sequences," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8, IEEE, 2007. (Cited on page 171.)

[326] W. R. Kunst-Wilson and R. B. Zajonc, "Affective discrimination of stimuli that cannot be recognized," *Science*, vol. 207, no. 4430, pp. 557–558, 1980. (Cited on page 176.)

[327] M. Josefsson, "Characterizations of trapezoids," in *Forum Geometricorum*, vol. 13, pp. 23–35, 2013. (Cited on pages 177 and 181.)

[328] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, pp. 2331–2338, IEEE, 2006. (Cited on page 177.)

[329] N. A. Dodgson, "Resampling radially captured images for perspectively correct stereoscopic display," in *Photonics West'98 Electronic Imaging*, pp. 100–110, International Society for Optics and Photonics, 1998. (Cited on page 177.)

[330] I. Dagan, M. C. Golumbic, and R. Y. Pinter, "Trapezoid graphs and their coloring," *Discrete Applied Mathematics*, vol. 21, no. 1, pp. 35–46, 1988. (Cited on page 179.)

[331] M. Hota, M. Pal, and T. K. Pal, "An efficient algorithm to generate all maximal independent sets on trapezoid graphs," *International journal of computer mathematics*, vol. 70, no. 4, pp. 587–599, 1999.

[332] C. Flotow, "On powers of m-trapezoid graphs," *Discrete Applied Mathematics*, vol. 63, no. 2, pp. 187–192, 1995.

[333] G. B. Mertzios and D. G. Corneil, "Vertex splitting and the recognition of trapezoid graphs," *Discrete Applied Mathematics*, vol. 159, no. 11, pp. 1131–1147, 2011.

[334] L. Droussent, "On a theorem of j. griffiths," *The American Mathematical Monthly*, vol. 54, no. 9, pp. 538–540, 1947. (Cited on page 179.)

[335] C. Bradley, "Three centroids created by a cyclic quadrilateral," *in Article: CJB/2011/141*, pp. 1–3, 2011. (Cited on page 183.)

[336] M. Josefsson, "Five proofs of an area characterization of rectangles," in *Forum Geometricorum*, vol. 13, pp. 17–21, 2013. (Cited on page 185.)

[337] M. F. Mammana, B. Micale, and M. Pennisi, "Properties of valtitudes and vaxes of a convex quadrilateral," in *Forum Geometricorum*, vol. 12, pp. 47–61, 2012. (Cited on page 185.)

[338] H. Fradi and J.-L. Dugelay, "Improved depth map estimation in stereo vision," in *IST/SPIE Electronic Imaging*, pp. 78631U–78631U–7, International Society for Optics and Photonics, 2011. (Cited on pages 187 and 188.)

[339] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 49–65, 2007. (Cited on page 188.)

[340] Y.-C. Fan and T.-C. Chi, "The novel non-hole-filling approach of depth image based rendering," pp. 325–328, IEEE, 2008. (Cited on page 188.)

[341] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Visual Communications and Image Processing 2003*,

pp. 573–582, International Society for Optics and Photonics, 2003. (Cited on page 189.)

[342] imatest, "image quality factors." Available from: http://www.imatest.com/docs/iqfactors/, 2014. (Cited on page 211.)

[343] E. Medic and M. Soltani, "Methods for characterization of digital, image-producing detectors within medical x-ray diagnostics," *University essay from Blekinge Tekniska Hogskola/Sektionen for Teknik (TEK)*, 2005. (Cited on page 212.)

[344] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "Objective metrics for quality of experience in stereoscopic images," pp. 3105–3108, IEEE, 2011. (Cited on page 213.)

colophon

This thesis was published at Brunel University, London. It is a documentation of the research conducted at the Centre for Media Communication Research. :

Abdulkadir.Audu@brunel.ac.uk

Final Version as of March 17, 2015 (classicthesis version 4.1).

DECLARATION

I declare that no portion of the work referred to in this thesis has been submitted in support of any application for another degree or qualification of any other University or Institution of learning.

London, January 2015

Abdulkadir Iyyaka Audu