3D Panoramic Imaging for Virtual Environment Construction

A Thesis Submitted to the University of Huddersfield in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

by

Duke Gledhill BSc (Hons)

School of Computing and Engineering

The University of Huddersfield

In Collaboration with Rotography Ltd

Sponsored by the EPSRC

May 12, 2009

ABSTRACT

The project is concerned with the development of algorithms for the creation of photo-realistic 3D virtual environments, overcoming problems in mosaicing, colour and lighting changes, correspondence search speed and correspondence errors due to lack of surface texture.

A number of related new algorithms have been investigated for image stitching, content based colour correction and efficient 3D surface reconstruction. All of the investigations were undertaken by using multiple views from normal digital cameras, web cameras and a "one-shot" panoramic system. In the process of 3D reconstruction a new interest points based mosaicing method, a new interest points based colour correction method, a new hybrid feature and area based correspondence constraint and a new structured light based 3D reconstruction method have been investigated.

The major contributions and results can be summarised as follows:

- A new interest point based image stitching method has been proposed and investigated. The robustness of interest points has been tested and evaluated. Interest points have been proved robust to changes in lighting, viewpoint, rotation and scale.

- A new interest point based method for colour correction has been proposed and investigated. The results of linear and linear plus affine colour transforms have proved more accurate than traditional diagonal transforms in accurately matching colours in panoramic images.

- A new structured light based method for correspondence point based 3D reconstruction has been proposed and investigated. The method has been proved to

increase the accuracy of the correspondence search for areas with low texture. Correspondence speed has also been increased with a new hybrid feature and area based correspondence search constraint.

- Based on the investigation, a software framework has been developed for image based 3D virtual environment construction. The GUI includes abilities for importing images, colour correction, mosaicing, 3D surface reconstruction, texture recovery and visualisation.

- 11 research papers have been published.

LIST OF PUBLICATIONS

1. M. Bingham, D. Taylor, D. Gledhill, Z. Xu, *"Integration of Real and Virtual Light Sources in Augmented Reality Worlds"*, 14th International Conference on Automation and Computing. Pacilantic International, ISBN 978-0955529320, pp 75-80. Brunel University, UK, September 2008.

2. G. Y. Tian, R. Lu, D. Gledhill, *"Surface Measurement Using Active Vision and Light Scattering"*, Optics and Lasers in Engineering, Vol. 45, Issue 1. Elsevier Science, ISSN 01438166, pp 131-139. 2007.

3. G. Y. Tian, D. Gledhill *"Visualisation Based Feedback Control for Multiple Sensor Fusion"*, 10th International Conference on Information Visualization (IV'06). IEEE Computer Society, ISSN 1550-6037, pp 553-556. London, UK, July 2006.

4. R. Lu, G. Y. Tian, D. Gledhill, S. Ward, *"Grinding Surface Roughness Measurement Based on the Co-occurance Matrix of Speckle Pattern Texture"*, Applied Optics, Vol. 45, Issue 35. Optical Society of America, pp 8839-8847. December 2006.

5. D. Gledhill, G. Y. Tian, D. Taylor, D. Clarke, *"Panoramic Imaging Based e-laboratory Construction"*, The Book of Advances in e-Engineering and Digital Enterprise Technology (e-ENGDET). Wiley, ISBN 978-1860584671, pp 589-600. August 2004.

6. G. Y. Tian, D. Gledhill, *"Structured Light Based Stereo Vision for Coordination of Multiple Robots"*, 1st International Conference on Informations in Control,

Automation and Robotics (ICINCO). Institute for Systems and Technologies of Information, Control and Communication, ISBN 972-8865-12-0, pp 158-161. Setubal, Portugal, August 2004.

7. C. Mitchell , G. Y. Tian, D. Gledhill, D. Taylor, *"Web-based Interactive 3D Visualisation for Business and Building Management"*, Proceedings of the 8th IASTED International Conference on Internet and Multimedia Systems and Applications (IMSA). ACTA Press, ISBN 0-88986-420-9. Kauai, Hawaii, USA, August 2004.

8. D. Gledhill, G. Y. Tian, D. Taylor, D. Clarke, *"3D Reconstruction of a Region of Interest Using Structured Light and Stereo Panoramic Images"*, 8th International Conference on Information Visualisation (IV'04). IEEE Computer Society, ISBN 0-7695-2177-0, pp 1007-1012. London, UK, July 2004.

9. D. Gledhill, G. Y. Tian, D. Taylor, D. Clarke, *"Panoramic Imaging - A Review"*, Computers and Graphics 27(3), ISSN 0097-8493, pp 435-445. 2003.

10. G. Y. Tian, D. Gledhill, D. Taylor, *"Comprehensive Interest Point Based Imaging Mosaic"*, Pattern Recognition Letters 24(9-10). ISSN 0167-8655, pp 1171-1179. 2003.

11. G. Y. Tian, D. Gledhill, D. Taylor, D. Clarke, *"Colour Correction for Panoramic Imaging"*, 6th International Conference on Information Visualisation (IV'02). IEEE Computer Society, ISBN 0-7695-1656-4, pp 483-488. London, UK, July 2002.

LIST OF FIGURES

2.1 A complete panorama 6
2.2 Kaidan Kiwi990 with Nikon Coolpix 990 camera 8
2.3 Diagram showing panoramic capture process for single camera where P_n = Photo, P_c = combined panorama and O = optical centre of the camera. 9
2.4 A 360° by 120° panoramic image 10
2.5 iMove SVS-2000 Camera System 11
2.6 Two 180° hemispheres that can be used to create an omni directional image 12
2.7 BeHere Video System uses a mirror to capture the environment . . . 13
2.8 The first true stereo camera with two lenses, built in 1849 14
2.9 A modern stereo camera setup 14
2.10 Altering the rotation point and using different parts of the images to create a stereo panorama 15
2.11 Original image, and position information provided by edge detector . 20
2.12 Left original image before the mosaicing 30
2.13 Right original images before the mosaicing 30
2.14 Two images stitched together (without blending) 31
2.15 Two images stitched together (with blending) 31
2.16 3D surface reconstruction 32
2.17 Stereo images and related disparity map 35
2.18 Epipolar geometry and epipolar plane 36
2.19 Triangulation to find a point in 3D from 2 2D images 38

3.1 The basic theory of the project 47
3.2 Showing the sub elements within each stage of the project 48

4.1 Stereovision system and panoramic imaging for 3D virtual environment 53
4.2 Showing the process of 3D panoramic imaging Where FOE is the Focus of Expansion and FOC is the Focus of Contraction, both parts of a panorama where either too much (FOE) or too little (FOC) depth is perceived. 54
4.3 The one-shot system 56
4.4 The multi-shot system, showing 4 of the 26 positions used 57
4.5 Stereo system using web cameras for stereo correspondence algorithm testing 57
4.6 Diagram showing multi-shot system camera orientations. 58
4.7 Screenshot of software framework. 59
4.8 A diagram of the software framework 60
4.9 Crops from consecutive images from a panorama sequence showing colour changes due to capture devices settings changes 62

5.1 Step 1 of the new approach of colour correction 64
5.2 Step 2 of new approach of colour correction 65
5.3 Linear transform matrix estimation and colour correction before image stitching and panorama 66
5.4 Histogram before and after equalisation 68
5.5 Overlapping area of the stitching images 71
5.6 Colour correction based on M or $M-1$ 71
5.7 A selection of images used in the experimental tests for the colour correction algorithms 89
5.8 Comparison of the speed for the colour correction pre-processing methods 90
5.9 Comparison of accuracy for the colour correction processing methods 91

5.10 Comparison of the average speed of the colour correction methods . . 92
5.11 Comparison of the differences in processing accuracy for different image variations 93
5.12 Comparison of the standard deviation in processing time for each pre-processing method 94
5.13 Comparison of the minimum, mean and maximum processing times for each pre-processing method 95
5.14 Different colour correction for panoramic imaging 97

6.1 Image mosaic processing block-diagram 99
6.2 Interest points across illumination and rotation 101
6.3 Rotation-invariant LBP 103
6.4 Two stitching images and their corresponding points 104
6.5 Stitched image without colour correction 105
6.6 The stitched images with colour correction 108
6.7 A selection of images used in the experimental tests for the feature matching algorithms 110
6.8 A comparison of the average accuracy (%) of the feature matching algorithms 115
6.9 A comparison of the average speed (seconds) of the feature matching algorithms 115
6.10 A comparison of the minimum accuracy (%) of the feature matching algorithms 116
6.11 A comparison of the minimum speed (seconds) of the feature matching algorithms 116
6.12 A comparison of the maximum accuracy (%) of the feature matching algorithms 117

6.13 A comparison of the maximum speed (seconds) of the feature matching algorithms . . . 117
6.14 A comparison of the standard deviation of accuracy (%) of the feature matching algorithms . . . 118
6.15 A comparison of the standard deviation of speed (seconds) of the feature matching algorithms . . . 118

7.1 An example of the structured light pattern to be projected onto low texture surface areas . . . 121
7.2 Showing the structured light pattern in an image (left) and after it has been removed (right) . . . 122
7.3 Showing the results of 3D depth calculations when no structured light pattern used . . . 123
7.4 Showing the results of 3D depth calculations when structured light pattern used . . . 124
7.5 Structured light pattern projection results showing fewer errors in the depth map where the structured light is projected (inside the red line) 125
7.6 Interest points and their correspondence . . . 125
7.7 Computing expense and search area . . . 126
7.8 Corresponding areas for constrained area based searching (Left image and right image) . . . 127
7.9 Process using interest points to increase the speed of processing the 2D data . . . 128
7.10 The stereo images and their disparity map, using area based method of correspondence searching . . . 128
7.11 Brick data showing Interest points and correspondence matches, and disparity map . . . 129

8.1 Embedded 3D virtual environment for Web-based interactive applications 131

LIST OF TABLES

5.1 Average (mean) values for all colour correction image tests 73
5.2 Minimum values for all colour correction image tests 74
5.3 Maximum values for all colour correction image tests 75
5.4 Standard Deviation for all colour correction image tests 76
5.5 Average (mean) values across the brightness variation colour correction test images 77
5.6 Minimum values across the brightness variation colour correction test images 78
5.7 Maximum values across the brightness variation colour correction test images 79
5.8 Standard Deviation of values across the brightness variation colour correction test images 80
5.9 Average (mean) values across the white balance variation colour correction test images 81
5.10 Minimum values across the white balance variation colour correction test images 82
5.11 Maximum values across the white balance variation colour correction test images 83
5.12 Standard Deviation of values across the white balance variation colour correction test images 84
5.13 Average (mean) values across the white balance and brightness variation colour correction test images 85

5.14 Minimum values across the white balance and brightness variation colour correction test images 86
5.15 Maximum values across the white balance and brightness variation colour correction test images 87
5.16 Standard Deviation values across the white balance and brightness variation colour correction test images 88

6.1 Average Interest Point Correspondence Results (SIFT and Proposed Method) 111
6.2 Minimum values for the interest point correspondence results (SIFT vs. Proposed Method) 112
6.3 Maximum values for the interest point correspondence results (SIFT vs. Proposed Method) 113
6.4 Standard deviation values for the interest point correspondence results (SIFT vs. Proposed Method) 114

ACKNOWLEDGEMENTS

I would like to thank the School of Computing and Engineering at the University of Huddersfield for granting me the opportunity to carry out this project. I would also like to thank Rotography Ltd and the EPSRC for co-funding the project.

I would like to express my gratitude to my director of studies, Prof. G. Y. Tian, who has guided and supported me throughout the project. My gratitude also goes to my second supervisor, Prof. D. Taylor for his help and support with the project.

I would also like to thank my parents for supporting and encouraging me throughout my education. They always pushed me to do better, and it doesn't get much better than a PhD!

The work was very enjoyable and I am very excited about the new Augmented Reality work currently underway. I would also like to keep working on the mosaicing using interest points, the Hugin application currently uses 2 different methods for automatic mosaicing. Developing a 3rd, faster, more accurate method would be very interesting.

TABLE OF CONTENTS

Abstract **i**

List of Publications **iii**

List of Figures **v**

List of Tables **ix**

Acknowledgements **xi**

Chapter 1: Introduction **1**

1.1 Motivation 1

1.2 Aims 3

1.3 Thesis Outline 4

Chapter 2: Literature Survey **5**

2.1 Introduction 5

2.2 Image Capture Systems 7

2.2.1 Single and Multi Camera-based Panoramic Systems 7

2.2.2 Omni Directional Imaging Systems 11

2.2.3 Stereo Cameras 13

2.3 Image Processing and Understanding 15

2.3.1 Noise Reduction 16

2.3.2 Radial Lens Distortion Correction 16

2.3.3 Camera Calibration 17

2.3.4 Feature Detection and Matching 18

2.3.5 SIFT 21
2.4 Colour Correction and Colour Constancy 22
2.5 Image Mosaicing and 3D Reconstruction 26
2.5.1 Mosaicing 27
2.5.2 3D Construction 31
2.5.3 Stereo Panoramas 40
2.5.4 Rendering and 3D visualisation 41
2.6 Discussions 42
2.6.1 Stereo Vision Based Panoramic Capture System 42
2.6.2 Deriving Capturing Conditions and Object Surface Characteristics 43
2.6.3 Image Understanding and 3D Reconstruction 43
2.6.4 Image Based Volumetric Rendering 44
2.6.5 Problems identified 45

Chapter 3: Theory and Project Issues 47
3.1 Project concept and major elements 47
3.2 Key project elements 48
3.2.1 Image capture 49
3.2.2 Pre-processing 49
3.2.3 3D reconstruction 50
3.2.4 Visualisation 50
3.3 Methodology 51

Chapter 4: Panoramic Stereo Imaging 52
4.1 Introduction 52
4.2 System Design 52
4.2.1 Hardware 55
4.2.2 Software 58

4.3 Correspondence 59
4.4 Content Based Colour Correction 61
4.5 3D Reconstruction 61

Chapter 5: Colour Correction **63**
5.1 Histogram Map Based Colour Correction 63
5.1.1 Histogram Equalisation 66
5.1.2 Histogram Mapping 67
5.1.3 Singular Value Decomposition 69
5.2 Colour Correction for Panoramic Imaging 70
5.3 Experimental Tests 72
5.4 Summary 96

Chapter 6: Image Mosaicing **98**
6.1 Introduction 98
6.2 4-step Interest Points Based Image Mosaic 98
6.2.1 Identification of Interest Points 99
6.2.2 Finding corresponding points from stitching images 101
6.2.3 Spatial and Spectral Transform Matrices 103
6.2.4 Image Mosaic with Smoothing 107
6.3 Experimental Tests 108
6.4 Summary 119

Chapter 7: 3D Surface Reconstruction **120**
7.1 Introduction 120
7.2 3D Surface Reconstruction Using Structured Light 120
7.3 Interest Point Based 3D Surface Reconstruction 122
7.4 Experimental Tests 126
7.5 Summary 127

Chapter 8: Conclusions and Further Work **130**

8.1 Panoramic Imaging for Virtual Environment Creation 130

8.2 Web-Based Interactive System Using 3D Panoramic Imaging 131

8.3 Video panoramic imaging 132

8.4 Using interest points to integrate the real and virtual world in computer games 132

8.5 Conclusions 132

Bibliography **135**

Chapter 1

INTRODUCTION

1.1 Motivation

Panoramic imaging is rapidly emerging as an important part of digital imaging and image processing, especially with the required equipment becoming much cheaper and more easily available. The most common and simplest way to produce panoramas is to use a consumer digital camera such as a Canon Digital IXUS [1] or a Pentax Optio S [2] which have a panorama assist mode. This means it is very simple to create panoramas as the camera directs where the next image should be taken from. Software is available for free [3] to be able to stitch the images together into a finished panorama. This level of panorama is at the cheaper and simpler end of the list of available options and needs approximately 20 photographs, depending on lens. Recently though, even mobile phones have panoramic assist modes for small, simple panoramas. Following these simple panoramic image capture systems more complex systems are available. There exist panoramic systems that can use two fisheye lenses [4], a single hyperboloid mirror [5, 6], or even a rotating camera designed to capture high dynamic range panoramas [7]. Many options already exist for panoramic image capture, from the inexpensive consumer system, to the expensive professional system.

3D reconstruction is another rapidly developing area in image processing. As with panoramic systems, off the shelf cameras are now utilizing simple stereo for viewing pictures as stereo images [2]. Using digital images for detailed 3D reconstruction is also becoming a consumer reality with software products like D Sculptor 2 from D Vision Works [8] or 3D SOM from Canon [9], although these still require the user to

do most of the work. Stereo vision is still largely a research area though, as many unresolved problems still exist. A significant problem is the correspondence problem, the most common and difficult for computer vision applications and algorithms. Correspondence describes the relationship of a feature or area between different images. Given a point or area in one image, the correspondence problem is that of finding the corresponding point or area in another image.

Recent research work has begun to combine the work of panoramic imaging and stereo imaging to create stereo panoramas. Commercially there is now a product from Autodesk which goes a long way toward this goal, although it is still an online process requiring time consuming user input. With more work in this field, it will ultimately mean the possibility of being able to create entire 3D environments from 2D panoramic images for use in virtual tourism, gaming, film, military, manufacturing and design. It will, in the future, be possible to enter a physical space, capture multiple panoramas at different locations and then create a 3D reconstruction of the scene, with little or no user involvement. Architects, film makers and games programmers could all take advantage of such technology to speed up existing methods of 3D construction.

Applications for a 3D panoramic system include military, tourism, games, film, manufacturing and design. As an example, a "drone" robot could scout a military zone with no risk to humans. The automated drone would capture panoramic data which would be processed in real time to generate a 3D virtual environment for soldiers or commanders to explore before entering the area. Another example might be being able to walk along a virtual beach to see a foreign resort, without leaving the comfort of your home. Games programmers could save time building complex models of famous buildings, just by visiting the building and capturing it in 3D. The success of a system with this capability would depend on accuracy, ease of use, speed and cost.

This work is partly sponsored by Rotography Ltd [10], a Halifax based panoramic imaging company. Rotography started working with panoramic imaging in 1997.

They are now working toward new technologies such as panoramic video and 3D panoramic imaging. They sponsored this project to further their work in panoramic technologies. They hope to gain useful insights into future technologies to help keep them ahead of the competition. The research is also sponsored by the EPSRC.

The University has a previous research record in the area of signal processing. Gui Yun Tian is a respected researcher in the areas of signal processing and colour correction. Duke Gledhill has a degree in Multimedia Technology from the University of Huddersfield and a personal, and expensive, interest in all technologies related to panoramic imaging.

1.2 Aims

The objective of this project was to develop and improve algorithms used in a 3D panoramic imaging system for fast construction of 3D environments from panoramic images. Experimental work was carried out to determine the most appropriate panoramic capture system for the purpose of 3D reconstruction. Experimental work was also undertaken to determine the most accurate methods of matching the panoramic images, and determining 3D information.

The objectives can be broken down as follows:

- To review current panoramic imaging techniques, from capture through to 3D reconstruction.

- To design a novel 3D panoramic imaging system, comprising capture, processing and visualisation.

- To apply advanced processing techniques for colour correction, feature detection and matching, and visualisation.

- To evaluate the test results and determine new strategies for overcoming any problems.

- To identify the key features of the system and applications for the system.

1.3 Thesis Outline

Chapter 1 discusses the background and objectives of the project. Chapter 2 discusses state-of-the-art research and techniques for the main areas of this research. Current panoramic capture systems and stereo capture systems are reviewed, including single and multi shot systems. The important components of panoramic imaging are covered including camera calibration, lens distortion, colour correction, correspondence for mosaicing, 3D reconstruction and visualisation. Chapter 3 discusses the research and how it has been separated into four main areas of research. Chapter 4 introduces a new 3D panoramic imaging system, including hardware and software. Chapter 5 introduces a new method of colour correction for panoramic imaging. Chapter 6 introduces a new image stitching method using interest points. Chapter 7 discusses a new interest points based method for 3D surface reconstruction from 2D images. Finally further work and conclusions are discussed.

Chapter 2

LITERATURE SURVEY

2.1 Introduction

As computers and cameras become cheaper, the use of digital images is becoming more prevalent. Panoramic imaging is an important part of this expanding use of digital equipment. The term panorama was first used by the painter Robert Barker to describe his panoramic paintings of Edinburgh. Surrounded by these paintings gave the impression of standing in another environment. Translating that to modern virtual tours, apart from being on a computer screen the feeling of immersion is still present. In its most general sense, a panorama is a single wide-angle image of the environment around the camera. Usually they completely surround the camera on the horizontal plane, and can be either approximately 120° in the vertical field of view, or 180° to create a complete sphere, but the term is often used to describe any wide angle representation of a scene. An example of a complete 360° cylindrical panorama is shown in Fig. 2.1 [10].

Panoramas are widely used in areas such as robotics, computer vision, surveillance and virtual reality. They have also been driven by commercial interests like entertainment, interactive TV, real estate and virtual tourism. In the past fifteen years, panoramic imaging systems have significantly progressed. Not only can specialists create and display panoramas, but also with a multitude of software available, some for free [11], everyone with a computer and camera is able to create panoramas. Even global corporations have embraced panoramas, for example Apple with their QuickTime software [12]. There are many ways to capture a panorama, either using a single

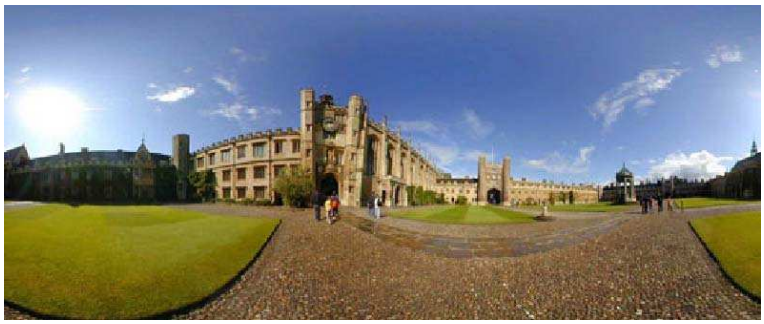

Figure 2.1: A complete panorama

camera mounted on a tripod and rotated about its optical centre, using a single omni directional camera, using multiple cameras facing in different directions, or using a stereo panoramic camera from which stereo information can be extracted. There is also another alternative to the stereo panorama for calculating stereo information. Instead of rotating the camera about its optical centre, the camera can be moved and used to capture multiple images of a scene from different viewpoints, from which stereo information can be calculated about the scene. This stereo information can then be used to create a 3D model of the scene, and arbitrary views can be computed. This information could also be used to build 3D models of large areas, for example a city might be reconstructed in this way. The process of the panoramic or stereo system can be simplified to four main stages:

1. Image Capturing System

2. Image Processing

3. Image Stitching or Stereo Processing

4. Rendering and 3D Visualisation

This chapter will review and compare some of the methods in the various stages of

3D panoramic imaging and identify research topics and applications. The rest of the chapter is organised as follows. Section 2.2 introduces image capture systems, section 2.3 discusses image processing and understanding, section 2.4 discusses colour correction, section 2.5 discusses image stitching and structure from stereo and finally 3D visualisation.

2.2 Image Capture Systems

As the technology used in digital cameras becomes cheaper, so do the cameras themselves. Modern digital cameras have high-resolution sensors, which capture a large amount of data about the scene being photographed. For panoramic imaging n number of images need to be captured, where n is dependent on the camera, lens, and the type of panorama required. Many panoramic image systems have been developed, from cheap off the shelf cameras that take a high number (e.g. >8) of sequential images, to omni directional systems that capture only one image.

The important parts of the imaging systems are:

- Resolution
- Field of view
- Capturing speed

2.2.1 Single and Multi Camera-based Panoramic Systems

Panoramic imaging from multiple images involves either a camera rotated about its optical axis (or nodal point), or using multiple cameras capturing different directions.

For a single camera rotated about its optical axis there are several components required

- Camera

- Lens
- Rotator
- Tripod

The camera and tripod can be relatively standard. The lens choice depends on the field of view required from each image. The lens choice determines field of view and therefore required number of images to capture the full panorama and also ensure enough overlap between images to enable correspondence matching and an accurate mosaic. The rotator is very important, it must enable the camera to rotate about its entrance pupil without flex or movement. An inaccurate rotator will cause parallax errors in the images and make mosaicing difficult or impossible.

From a commercial perspective, all of the above can be acquired easily on the open market. Panoramic imaging systems can start with relatively standard, off the shelf cameras [13, 14], panorama tripod mounts [15, 16] and software [11, 17, 18]. A commercial example might include a Nikon 990 [13] mounted on a Kiwi990 (Fig. 2.2) tripod mount from Kaidan [15], and using Pixaround [17] software to stitch the images. This example system would require several images for an accurate panorama, by altering the lens (e.g. a Nikon 8mm FC-E8), fewer images need capturing.

Figure 2.2: Kaidan Kiwi990 with Nikon Coolpix 990 camera

Once a multi-shot single camera system like the example above has been setup

the panorama is captured by turning the camera on the tripod about the optical centre of the camera (Fig. 2.3) and capturing an image at chosen points. Important considerations include overlap, movement of objects in the scene near the overlap regions and control of camera settings, such as white balance and exposure. Consistent white balance and exposure are important for smooth blending of images.

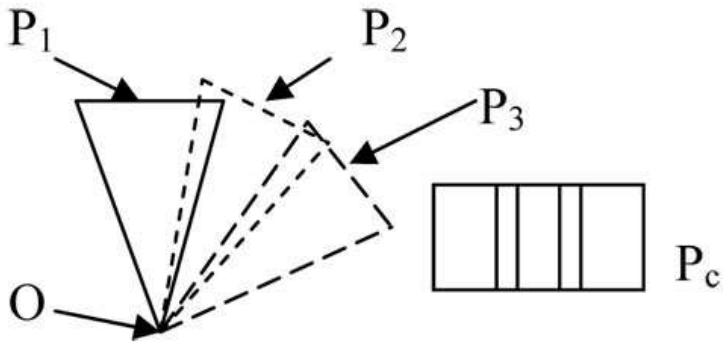

Figure 2.3: Diagram showing panoramic capture process for single camera where P_n = Photo, P_c = combined panorama and O = optical centre of the camera.

The main disadvantage of this multi-shot single camera system is that dynamic scenes can cause many problems, for best results static scenes are required. In dynamic scenes, motion may be present at the overlap of the images, and create erroneous information in the final panorama. Some software uses masking to attempt to compensate for differences in overlap regions. However, this is an off-line approach with varying degrees of success, for example in very dynamic scenes like a crowd of people it is difficult to acheive an accurate mosaic. Generally though multi-shot single camera systems give a much higher resolution image than the omni directional systems, for example a Canon 20D with EF-s 10-22mm lens using 13 shots creates a 49 mega pixel spherical panoramic image. An omni directional image from the same camera but with a 0-360 one-shot mirror only creates a 0.9 mega pixel cylindrical panoramic image. Multi-shot systems require the camera settings between shots to

be kept unchanged. For example changes in white balance and exposure between captures will result in more difficulties for mosaicing and blending images for a seemless panorama.

Advantages of this process are that it can be relatively cheap and easy to use, therefore it has become popular with many photographers. Also the fact that almost any camera can be used on just about any tripod (and rotator) and it makes this type of panorama even more appealing. Some modern cameras even come with panorama assist modes to help in the capture process (e.g. Canon Digital IXUS and Nikon P series cameras).

Figure 2.4: A 360° by 120° panoramic image

An alternative to using a single camera on a tripod is to use multiple cameras mounted together, each pointing in a different direction. For example a system with 6 cameras, each spaced 60° apart. This system has the advantage of being able to capture the whole scene in a single shot, and is able to keep the high resolution. This system could also be used to capture video panoramas. The disadvantage of this system is the number of cameras; mean time to failure for the system will be significantly lower. A commercial example of this type of system is the iMove SVS-2000 system [19] (Fig. 2.5). The iMove system uses 6 cameras, 4 on the horizontal plane, spaced 90° apart and 1 up and 1 down. This system is able to capture video rate spherical panoramas. Unfortunately because panoramas require images taken from a central nodal point the cameras have to be as close together as possible. This

translates into small cameras and with current technology relatively low resolution. As sensor technology improves higher resolution solutions will be possible. Only 4 cameras for 180° coverage also requires very wide angle lenses and the issues that exist with them, e.g. distortion and resolving power.

Figure 2.5: iMove SVS-2000 Camera System

2.2.2 Omni Directional Imaging Systems

Omni directional image sensors, first proposed in 1970 [20], capture the whole environment in one single image and have the advantage of speed of capture without the need for mosaicing. Panoramic video is also a possibility with this type of system. It also means that the scene does not have to be static, as any motion will still be captured in the same image. In an omni directional camera system there must be a single effective viewpoint if pure perspective images are to be calculated and 3D information extracted. Omni directional sensors either incorporate lenses or mirrors. Some commercial systems use fisheye lenses [4] (Fig. 2.6 [10]) and others use a variety of lens/mirror configurations (Fig. 2.7). Fisheye lenses are generally quite large and typically expensive. They have a very short focal length, which produces a hemispherical image, often with an approx. 180° field of view. There is a significant amount

of lens distortion with fisheye lenses which takes processing power to correct, making real time systems a more difficult goal. The low resolution and high lens distortion of the fisheye lens means that accurate distortion free perspective views cannot be calculated from images captured with a fisheye lens. Catadioptric imaging systems use a reflecting surface to enhance the field of view. Examples of Catadioptric systems can be found in [21] and commercial examples include the 0-360 one-shot system used in this project [6]. Svoboda *et al* describe Spherical, Hyperbolic and Parabolic mirror systems and derive information making them useful for stereo panoramic imaging [21]. However, omni directional systems generally suffer from lower resolution, something that will become less of an issue as higher resolution sensors are developed.

Figure 2.6: Two 180° hemispheres that can be used to create an omni directional image

A common problem with both single camera systems and multi camera or omni directional systems is the dynamic range. Dynamic range in photography refers to the limits of the luminance range that a camera can capture. The dynamic range of current cameras is relatively small compared to the human vision system. With the increase in the field of view there is generally an increase in the range, for example including both the sun and shadows in an outdoor panorama. Aggarwal and Ahuja [22] have developed a method to increase the dynamic range of the systems to produce better quality images. They propose the use of a mask with 50% normal and 50% covered by a neutral density filter. This ensures when rotating the camera both

Figure 2.7: BeHere Video System uses a mirror to capture the environment

high key and low key images are captured for stitching. This method could not be used for the multi-camera and omnidirectional systems. For omnidirectional sensors bracketing the capture settings of the camera and post processing to create a high dynamic range image can be very effective. As new sensors are developed, higher dynamic range images will be possible.

2.2.3 Stereo Cameras

Stereo vision consists of two images of a scene taken at different viewpoints. What our brains see as depth is called the disparity, which is the angular difference in viewing directions of each scene point between the two images. Disparity is a function of the points depth and the distance between eyes/cameras (baseline). Stereo cameras were invented around 1839 to take two photographs seperated by a short distance, usually similar to the human eye seperation. When viewed using a special viewer, so that each eye saw the correct image only, depth perception was possible. See Fig. 2.8 for an example of the first true stereo camera and a more modern stereo setup shown in Fig. 2.9. Many modern stereo cameras are available but need a viewer, like special goggles, to see the two individual images correctly. Using the function of

disparity and accurate correspondence information, computers are able to calculate depth information from a set of stereo images to produce depth maps of the scene.

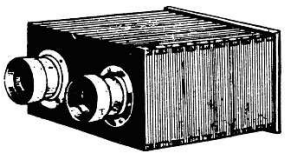

Figure 2.8: The first true stereo camera with two lenses, built in 1849

Figure 2.9: A modern stereo camera setup

For stereo panoramic image capture systems several options have been proposed. Huang and Hung [23] proposed a system using two cameras. Their PSI-II system is able to correct for problems such as epipolar alignment. Schum and Szeliski [24] also proposed a stereo panoramic system by offsetting the rotation point of the camera and using multi-perspective panoramas. Other researchers, for example Peleg and Ben-Ezra [25] and Ishiguro *et al* [26] have proposed single camera stereo panoramic systems. Peleg and Ben-Ezra propose rotating the camera about a point which is offset from the nodal point, actually behind it. Once the camera has been rotated the image is split into bands, for the left and right parts of the images. These left and

right components are mosaiced to create two panoramas, one for each eye, illustrated in Fig. 2.10.

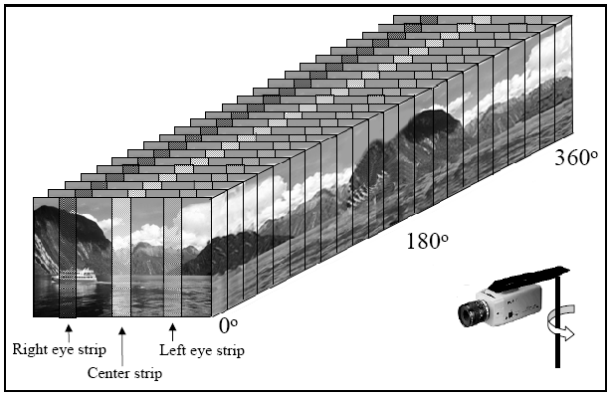

Figure 2.10: Altering the rotation point and using different parts of the images to create a stereo panorama

2.3 Image Processing and Understanding

The image acquired in the image capture system must now be processed. The processing is required because the computer does not 'know' what the images are, and what information they contain, apart from the pixel data. Once the images are processed they can be used in a panoramic or stereo system. There are many stages to image processing, some which are not used in both stereo and panoramic imaging. For 3D reconstruction and panoramic imaging the image processing can be defined in several stages:

- Noise Reduction
- Radial Lens Distortion Correction
- Camera Calibration

- Feature Detection and Matching

2.3.1 Noise Reduction

Noise is an inevitable part of image processing, particularly when using cheap off the shelf systems. In recent years as the market is driven by megapixels noise has become more of an issue for commercially available camera systems. Photosites are getting smaller and the sensors are staying the same size. Noise can affect the location of features, and therefore the correspondence problem. The greater the noise, the greater the results will be affected. To minimise the effects of noise in the calculation of the corresponding points, a Gaussian blur can be applied to the images, this makes the effects of noise less significant [27], but also reduces contrast making feature point detection less reliable. Tensor voting, presented by Medioni et al [28], is a novel methodology for the robust inference of features from noisy data. This method does not need the noise to be 'blurred'.

2.3.2 Radial Lens Distortion Correction

Most work with 3D assumes a pinhole camera model. A pinhole camera projects the world onto the image plane linearly if expressed in terms of projective geometry. However, optical systems in commercial cameras are not linear, so the camera is not linear. The camera model can be described as linear with non-linear distortions. Often the non-linear distortions can be ignored if very small and less accurate information is acceptable. If accurate information is required then the non-linear distortions need to be modelled and corrected. Lens correction is required where cheaper or extreme wide angle lenses are used, particularly in the off-the-shelf systems, where radial lens distortion is more significant. These radial lens distortion effects will mean that the pinhole camera model is not satisfied. Lens distortion correction remaps the distorted image to a distortion free image, which then satisfies the pinhole camera model [29].

The simplest perspective projection of a pinhole camera onto an image plane can be expressed as

$$\begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.1}$$

where the matrix describing the mapping is called the camera projection matrix P. In Eq. 2.1 the camera projection matrix is the simplest possible case and only contains information about the focal distance f.

If Eq. 2.1 is simplified to

$$zm = PM \tag{2.2}$$

then $M = (x, y, z, 1)^T$ are the homogeneous coordinates of the 3D point and $m = \left(\frac{fx}{z}, \frac{fy}{z}, 1\right)^T$ are the homogeneous coordinates of the image point.

2.3.3 Camera Calibration

Camera calibration provides information about the intrinsic (focal length, aspect ratio, image centre and radial distortion coefficient) and extrinsic (rotation matrix and translation vector) parameters of the camera. The intrinsic information provides data for estimating the camera model with more accuracy, rather than assuming the pinhole model. Camera calibration is a necessary stage in 3D computer vision if 3D information is to be extracted from the images. The camera information is also useful for computing viewpoint changes in 3D environments and panoramic image navigation. There are two types of camera calibration, photogrammetric calibration and self-calibration.

Photogrammetric calibration involves capturing a known calibrated object from multiple viewpoints. Photogrammetric calibration is very efficient, but needs expen-

sive calibration equipment and a complex setup. Zhang [30] and Heikkila [31] have accomplished useful work in this area. Zhang proposes a flexible technique which requires the camera observe a planar pattern at two or more orientations. Either the camera or the pattern can move, no other information is required.

Self-calibration uses no known object or pattern. It instead uses multiple captures of a static scene to calculate calibration information. McMillan [32] and Luong and Faugeras [33] have accomplished a lot of work in this area. Luong and Faugeras propose a new Fundamental Matrix which is a simplification of the Essential Matrix to provide epipolar geometry information. The Essential Matrix requires a calibrated camera and intrinsic parameters. The Fundamental Matrix describes the correspondence in more general terms.

In this project camera calibration is achieved using the Camera Calibration toolbox from Strobl *et al* at the Institute of Robotics and Mechatronics [34] which is based on the work of Zhang [30].

2.3.4 Feature Detection and Matching

To extract useful information from the scene, features need to be detected in the images and then matched with corresponding features in the other images. The correspondence problem is accepted as one of the most difficult and ongoing problems in image processing. Many applications of 2D image registration, or matching, are covered in Odone and Fusiello's paper [35] and readers are referred to the paper by Brown [36] for an in depth review of image registration techniques. There are two main types of detecting and matching features in images, area based methods and feature based methods.

Area based techniques use correlation of intensity patterns of a pixel with the intensity pattern around the corresponding pixel in another image. First a point of interest is chosen in one image. Then cross-correlation is used to search for the corresponding pixel in the other image. Decisions about the size of the search window

and the location and size of the search region need to be made. Some area based methods use adaptive window sizing. Area based methods suffer because they use the intensity values at each pixel directly, and are therefore sensitive to changes in viewing position, absolute intensity, contrast and illumination. Occlusions can also give erroneous correspondence information. Zitnick and Kanade [37] use a co-operative stereo algorithm for stereo matching applying uniqueness and continuity constraints to derive occlusion information.

Feature based methods are based on intensities in the images (e.g. edges, corners), rather than image intensities (e.g. pixel colour) themselves. There are two features that are most commonly used, edges and interest points.

Edge detectors attempt to recover the discontinuities in the photometric, geometrical and physical characteristics of objects in the images. This information creates variations in the grey-level image. There are three steps to edge detection. The first step is noise smoothing to suppress as much of the image noise as possible, without destroying the true edges. The second step is edge enhancement, which means applying a filter designed to be large at edge pixels and small elsewhere. The third step is edge localisation, which means deciding which local maxima in the filters output are edges and which are just caused by noise. Nonmaximum suppression to thin wide edges and thresholding can be used for edge localisation. There are three main ways to use the variation information, discontinuities (step edges), local extrema (line edges) and the 2D features formed where at least two edges meet [38]. Fig. 2.11 [30] shows an example of an original image and the position information provided by an edge detector. There are tens of edge detectors in the vision community, but they all produce similar results. Common edge detectors include Canny [39] and Sobel. Canny proposed the original edge detector in 1986, but it is still state-of-the-art today and still the most used.

Bao and Xu [40] use edge-preserving visual perception modelling to build mosaics. Kim et al [41] use edges in their graph matching algorithm and use camera

focal length to improve stitching reliability, which finds corresponding edges for the matching. Edge detection is used in Bourque's et al [42] work for determining interesting locations from where a robot should capture a spherical panorama.

Figure 2.11: Original image, and position information provided by edge detector

Interest point detectors attempt to find locations in the image where the signal changes two-dimensionally, locations such as corners, T-junctions or significant changes in texture. Interest points for image matching can be traced back to the interest operator used by Moravec [43] for automatic rover obstacle avoidance. Later Harris and Stephens [44] improved the Moravec detector to make it more repeatable. Many interest point detectors now exist, as reviewed by Schmid et al [45]. Schmid et al showed that the Harris corner detector was the most consistent when testing repeatability and information content of interest point detectors. Interest points, as local invariant features [46], are robust to changes in lighting, viewpoint, rotation, translation and scale [47] and are used in many areas including image retrieval and indexing [48], object recognition [49] and mosaicing [46] to cite a few. More recently a new approach to low level image processing has been introduced. The SUSAN principle developed by Smith and Brady [50] uses non-linear filtering to define parts of an image which closely relate to an individual pixel, each individual pixel now having an associated local image region of similar brightness to that pixel. The new SUSAN feature detectors are based on the minimisation of this local image region.

2.3.5 SIFT

Scale-invariant feature transform (SIFT), first proposed by Lowe in [49] and updated in [51] uses interest points (or keypoints) that have proved invariant to scale and rotation and robust to changes in viewpoint and illumination. The SIFT algorithm will be used for testing proposed methods against so is described here in more detail.

There are 4 main stages to the SIFT algorithm:

- Scale-space extrema detection
- Keypoint localisation
- Orientation assignment
- Keypoint descriptor

Scale-space extrema detection

Candidate keypoints are detected using a Gaussian based cascading filtering approach. Candidate locations are compared across different scale-spaces [52] to find points that are invariant to scale changes. Convolution of a variable scale Gaussian filter with an image produces multiple scales of an image. Adjacent scales are compared using a difference-of-Gaussian function. This subtracts one scale space image from another. Once multiple d-o-G images are computed, possible delegates for keypoint selection are compared with their 8 neighbours, and 9 neighbours in the scale above and below. Stable keypoints are those which are either larger or smaller than all of the neighbours.

Keypoint localisation

Keypoint selections need to be filtered for points that have low contrast or are poorly located along an edge. A 3D quadratic function is fitted to the local sample point to determine the interpolated location of the maximum. A threshold is applied to

reject points with a low contrast. Next poorly located edge points are rejected. Edges produce a strong response in the difference-of-Gaussian function. A poor edge point can be defined as having a large principle curvature across the edge, but a small one in the perpendicular direction. By applying a threshold to the ratio of the principle curves across and perpendicular to the edge, poor edge points can also be rejected.

Orientation assignment

Invariance to rotation can be acheived by assigning a consistent orientation to each keypoint. An orientation histogram is formed from the gradient orientations of keypoints. Peaks in the orientation histogram correspond to dominant directions of local gradients. The highest peak on the histogram is used to determine orientation information. If there exists another peak within 80% of the highest then multiple keypoints are created, each with a different orientation.

Keypoint descriptor

The previous stages have provided a local 2D coordinate system in which to describe the local image region using keypoints with invariance to scale and orientation. This final stage computes a descriptor to provide invariance to illumination and viewpoint changes. The descriptor is formed from a vector based on orientation histograms taken from the surrounding pixels around a keypoint. The feature vector is also normalised to reduce the effects of global contrast and brightness changes, making the descriptor robust to lighting changes.

2.4 Colour Correction and Colour Constancy

Colour changes are a common problem in panoramic imaging due to changes in lighting, viewpoint, or even uncontrollable device settings which all have an effect on the final captured image. The basic method of colour correction used currently is the

diagonal model, shown in Eq. 2.4 which was adapted by Ives [53] and based on the theories of Von Kries.

The digital camera records large shifts in image colours under different illuminations. However, a human observer viewing each scene will be able to discount the colour of the illumination and perceive the colours in each scene as the same. This property of compensating for illumination is called colour constancy. Colour constancy is a subconscious colour correction that all humans have. To overcome the problem of colour distortion when capturing images under different lighting or different camera properties, it is necessary to understand the principle of colour variation between the two images. The machine colour constancy problem can be defined as follows. First, choose some illumination as the standard, or canonical, illumination. The choice of canonical illumination matters little so long as it is not unusual. Then consider the 3 band RGB image obtained by any standard colour camera of a scene under some other, unknown illumination. The machine colour constancy problem requires converting the RGB at every pixel to be what it would have been had the same scene been illuminated by the canonical illumination. In this way, all the RGB values in the image of the scene measured under the unknown illumination are converted to standardized RGB descriptors relative to the canonical illuminant. Once these standardized descriptors have been obtained, they can be used for object recognition or for creating an image of the scene as it would appear under some other illuminant. The key to solving this problem is discovering the colour of the unknown illumination. Two well-known colour constancy methods, which work under limited circumstances, are the grey world algorithm and the white patch algorithm. The grey world algorithm assumes that the average of all colours in an image is grey, i.e. the red, green and blue components of the average colour are equal. The amount the image average differs from grey determines the illuminant RGB. The white patch algorithm, which is at the heart of many of the various retinex [54] algorithms, presumes that in every image there will be some surface or surfaces such that there will be a point or points

of maximal reflectance for each of the R, G, and B bands.

A common approach to colour constancy is the use of the estimation illumination to correct the images to a canonical light. Finlayson et al [55] suggested that if a transform is linear, a diagonal model might be sufficient to model the colour transform. Generally colour cameras are tri-chromatic, which means in a colour image, each pixel is a 3 vector, one component per sensor channel and works independently. However, with increasing colour fidelity, more accurate transforms will be required [56].

Different linear colour transforms, where the colour variation may be caused by lighting, changes in viewpoint or capturing devices, are discussed as follows. The transform Matrix M across images I_1 and I_2 can be represented as

$$I_1 * M = I_2 \tag{2.3}$$

1) Diagonal model

$$M = \begin{bmatrix} \alpha & & \\ & \beta & \\ & & \gamma \end{bmatrix} \tag{2.4}$$

$$\alpha = \frac{mean(R_2)}{mean(R_1)} \tag{2.5}$$

Where R is the red channel image intensity values in the two images. β and γ are similar for green and blue channels.

General features:

- Simple, not accurate enough in some cases
- Based on greyworld principle and does not need matching geometric pixels in the 2 images

2) Diagonal model plus affine transform

$$M = \begin{bmatrix} \alpha & & \alpha_1 \\ & \beta & \beta_1 \\ & \gamma & \gamma_1 \end{bmatrix} \tag{2.6}$$

α, β and γ will be the same as diagonal model. The offset can be obtained from polyfit in the individual channels.

General features:

- More accurate than diagonal model
- Two images require the same corresponding pixels

3) Linear model

$$M = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{2.7}$$

Where M can be computed by

$$M = \left[I_1^T I_1\right]^{-1} I_1^T I_2 \tag{2.8}$$

Where

$$I_1, I_2$$

is an $[n, 3]$ matrix and n is the number of pixels in the images.

General features:

- Good accuracy
- Need the same corresponding pixels in both images
- Computationally expensive

4) Linear model with affine

$$M = \begin{bmatrix} a & b & c & a_1 \\ d & e & f & e_1 \\ g & h & i & i_1 \end{bmatrix} \tag{2.9}$$

In addition to equation 2.8, the offset can be obtained by

$$\begin{bmatrix} a_1 \\ e_1 \\ i_i \end{bmatrix} = \begin{bmatrix} mean(R_2) \\ mean(G_2) \\ mean(B_2) \end{bmatrix} - \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \times \begin{bmatrix} mean(R_1) \\ mean(G_1) \\ mean(B_1) \end{bmatrix} \tag{2.10}$$

The general features are the same as the linear model.

As described above, to obtain better colour correction, more parameters in linear transforms will be used. We will use and compare diagonal model plus affine transform, linear model and linear model plus affine for colour correction in image correction for panoramic imaging.

2.5 Image Mosaicing and 3D Reconstruction

For panoramic imaging the images that have gone through image processing need to be mosaiced, or stitched, together.

For 3D reconstruction, the stereo images are used to calculate the disparity between the image data.

For both of these processes, accurate correspondence information is essential. The correspondence information is used to generate a homographic matrix, which remaps the images onto the same plane or is used in the stereo process for calculating the distance between corresponding pixels, to produce a disparity map.

2.5.1 Mosaicing

The creation of panoramic images and mosaics from a video sequence or a collection of images has attracted tremendous attention from researchers and commercial practitioners alike. Most systems for creating panoramas require the use of special fixtures (e.g. tripods and rotators) for precisely controlled image capture [57] and are transformed by their geometrical mapping [58].

In the last few years general interest in mosaicing has proliferated the vision and graphics community because of the range of possible applications i.e. teleconferencing, e-commerce, reconstruction of virtual environments and games. Szeliski et al have presented techniques for automatically deriving realistic 2D scenes and 3D texture mapping models from video sequences with applications in virtual environments [58, 59, 60]. The principal task in image mosaicing is finding the corresponding points and their transforms from the source images, especially when the stitching images have been produced in different capture conditions with different viewpoints and capture devices. One of the approaches is to compute eigenimage features using principal component analysis (PCA) for finding corresponding areas [61]. Another approach applies wavelet-based edge preserving for finding corresponding image features [40]. However, edge-based features are not robust to viewpoint changes. Also recent progress on pattern recognition, based on local features, invariant features in particular, has been used in many applications [49]. Interest points have been extensively investigated for object recognition and content-based image retrieval in computer vision [49, 62, 63] and corner-based image mosaicing has been presented with reasonably good results [46]. The SIFT algorithm has proven very robust and popular as a correspondence search and matching method for panoramic imaging and object recognition.

A mosaic is a collection of images and the transformations that relate them. In the case of a collection of images of a planar scene, taken from different points of view,

or a collection of images of a 3D scene taken from the same point of view, the only difference between the images is a rotation around the optical center of the camera and the transformation between the images is a linear transformation of projective space, called a collineation or a homography [64]. Previous work simplifies 2D image mosaics by the homography estimation [33]. Two cases have been investigated. The first is when the homography is mainly a translation and the rotation around the optical axis and zooming are small. The second is the general case where the rotation around the optical axis and zooming are large. Some efficient methods have been developed to handle the first case. For example, if the overlap of the images is very large, it has been shown that a non linear criterion minimization using the Levenberg-Marquardt method yields very good results [58], but it is very sensitive to the local minima and computationally expensive. A corner-based method to compute the homography between two images with small overlap (around 50%) and arbitrary rotation around the optical axis has been presented and used to build a 2D mosaic from a set of images [46, 65]. For 3D object images, a precise imaging mosaic has exploited non-linear transforms, for example the quadratic transform [65, 66]. This is the second-order Taylor expansion of the general interframe mapping function where the usual affine transformation model is the first-order expansion.

As described above, previous work to date mainly concentrated on geometric transformations for image mosaics. With the wide use of colour imaging, spectral transformation is also becoming important, particularly for dealing with image mosaics under various conditions. Colour image transformation and correction for illumination can apply linear models [67, 56]. The diagonal model plus affine transform will be used in this work.

New developments in electronic imaging and computer technologies now enable the generation, processing, communication and display of complex and rich images. Imaging has been developing rapidly over recent years and there is considerable scope for innovation and refinement. The constant drive to accumulate multiple sensor

devices into larger, faster, more efficient arrays stimulates technological developments in both the hardware and software aspects of imaging techniques. Improvements in identification and recognition have long been recognised as the primary goals of intelligent systems. The motivation for this work is the many potential applications in teleconferencing, construction of image-based virtual environments, e-commerce, medical imaging and panoramic imaging.

For panoramic stitching or mosaicing, the features that are found in the previous methods are then used to calculate the homograph between the two images. The homograph is a mapping between two perspective images of a planar surface in a scene [68, 69]. In effect this means that two images of a scene are related by a homograph, typically a 3x3 matrix (see eq. 2.11). This homograph can be used to re-map images onto different planes.

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} \\ H_{2,1} & H_{2,2} & H_{2,3} \\ H_{3,1} & H_{3,2} & H_{3,3} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.11}$$

In an automated system, there are many registration points and some of these points will contain false information, i.e. incorrect registration. Deciding which of these registration points to use to calculate the homograph is a difficult proposition. By using all the points to create multiple homographic matrices and then filtering them is one solution to finding the correct homographic matrix. There are a few methods to find the 'best fit'. One common practice is to use the least squares method. Another more recent method by Kanatani et al [68] uses a theoretical accuracy bound based on a mathematical model of image noise, and then a technique called renormalisation that theoretically attains the accuracy bound in the first order.

Once the optimal homographic matrix has been calculated the homograph information is then used to re-map the images onto the same plane, and place them in the correct positions so that the individual images create a single panoramic image

such as the one in Fig. 2.14. The image in Fig. 2.14 used a homographic matrix to align the images from Fig. 2.12 and Fig. 2.13 correctly after the 'best fit' matrix was computed from feature points in the images. Then colour correction algorithms were applied to create the final image, shown in 2.15. Mann and Picard [70] use video sequences and mosaicing to build high resolution images, and Kourogi et al [71] have created a real time mosaicing system using video sequences. Zhu et al [72] use video mosaics for stereo imaging.

Figure 2.12: Left original image before the mosaicing

Figure 2.13: Right original images before the mosaicing

Burt and Adelson [73] use a multi resolution spline technique for combining two

Figure 2.14: Two images stitched together (without blending)

Figure 2.15: Two images stitched together (with blending)

or more images into a larger mosaic. In this technique the images are used as a set of band-pass filtered component images, instead of using the whole image data. Onoe et al [74] use omni directional video streams to create real-time view-dependent perspective images, which lends itself to real-time telepresence systems.

2.5.2 3D Construction

The problem is that of 3D reconstruction, which is determining the 3D position of points in a scene from a 2D representation. For example when a 2D image, e.g. digital photograph, is captured the 3D scene is projected onto the 2D plane. 3D reconstruc-

tion attempts to recover the 3D scene from this 2D version. It is the well-known ill-posed computer vision problem. Particularly, the reconstruction of a dynamic, complex 3D scene from multiple images is an old and challenging problem. Early work by Longuet-Higgins [75] into deriving 3D information from 8 points led the way for 3D reconstruction and his work is still the basis for many of todays algorithms. Numerous studies have been conducted on various aspects of this general problem, such as the recovery of the epipolar geometry between two stereo images, the calibration of multiple camera views, stereo reconstruction by solving the correspondence problem, the modelling of the occlusions, the fusion of stereo and motion, and the fusion of multiple images by lighting variation [76]. Of course the most accurate method of 3D reconstruction is to use Laser measurement. This can be expensive and where people are included in the panoramic scene a risk to health exists. Yu *et al* [77] used a laser mounted on a van to produce accurate models of road surfaces and a video camera to produce the texture.

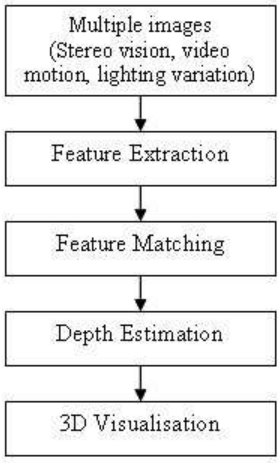

Figure 2.16: 3D surface reconstruction

Algorithms for surface reconstruction from different cues, such as stereo, shading, shadows, texture and spectral properties can be summarised in Fig. 2.16. Integrating

information from different sources such as stereo vision, video sequences or multiple images under different lighting are input for feature extraction. The feature extraction will obtain point, edge, area, colour, texture features etc for matching corresponding features in different image resources. Then depth estimation will calculate the depth information for 3D data and visualisation. Many approaches have been investigated in the last few decades. More details are discussed in the following paragraphs.

Using the correspondence information, a disparity map can be computed, as illustrated in Fig. 2.17 [78]. The disparity is the distance between corresponding pixels, and can be used to triangulate the 3D position of the pixels in the image. One of the biggest problems in stereo imaging, apart from correspondence, is the problem of occlusions. Stereo imaging uses multiple views of the same scene, either with the same camera from different viewpoints, or with multiple cameras. There are many ways of calculating the disparity from stereo images, including using structured light, geometry based approaches, image based approaches and parallax based approaches. Dhond and Aggarwal [76] reviewed the state-of-the-art techniques up until 1989 in their paper, which this section continues. Park and Inoue [79] use multiple cameras to capture information, and use the information in a hierarchical depth mapping system to overcome the occlusion problem. Hamden et al [80] use a trinocular vision system together with structured light to create a fast 3D object reconstruction system. Valkenburg and McIvor [81] use a single camera with a projector which projects a stripe pattern onto the object which the camera captures.

Pilu [82] suggested using Singular Value Decomposition of an approximate correspondence strength matrix, which gives very good results. Tzovaras et al [27] use dynamic programming to calculate the depth map. Tsai and Katsaggelos [83] presented a divide-and-conquer approach for dense disparity estimation. Aquigiar and Moura [29] use a surface-based factorisation method to recover 3D structure. Taylor and Jelinek [84] construct linearly parameterized models from single images with a camera of unknown focal length. Once the structure of the scene is known arbitrary

views of the scene can be calculated. Izquierdo and Kruse [85] use stereo imaging for calculating arbitrary views of the extracted models. Lee et al [86] use a similar approach to stereo, using a front, back and side image, to create human models from photographs. The system uses interest points to determine the silhouette of the human. Sietz and Dyer [87] use two images in a similar way to stereo vision, but they use them to generate physically valid synthetic views of the scene, in effect viewing the scene from a different viewpoint, not captured in the original images. Debevec et al [88] approach the model generation technique from a different angle. They use a manual photogrammetric method for the initial simple model, and then use a model based stereo algorithm, which recovers how the real scene differs from the model. This technique is limited to architectural models though, because it uses constraints that are characteristic of architectural scenes. Oh et al [89] use a single image, with much user input to generate 3D models. The system relies on the users input for assigning layers, and depths. Aliaga and Carlbom [90] use multiple panoramas to create a new walkthrough of a large environment. Their system captures the panoramas and uses them to plot a series of interlocking "walkways", which can be used to generate the new walkthroughs.

Correspondence is the most important process in 3D surface reconstruction. The correspondence matching includes point matching, scan line matching, edge with region matching and segment area matching from images under different capturing conditions such as different viewpoints (structure from stereo) [91, 92, 93, 94], movement (structure from motion) [95, 96, 72] and different lighting [97, 98]. Stereo vision infers 3D structure from two or more images taken from different viewpoints. Two or three cameras are often used, which are binocular vision or trinocular vision. More than two cameras used for stereo vision can improve the 3D reconstruction accuracy and efficiency [91, 93]. To estimate depth from a pair of stereo images, two main problems need to be solved. They are the correspondence problem and reconstruction problem. The correspondence problem finds the corresponding item in the right

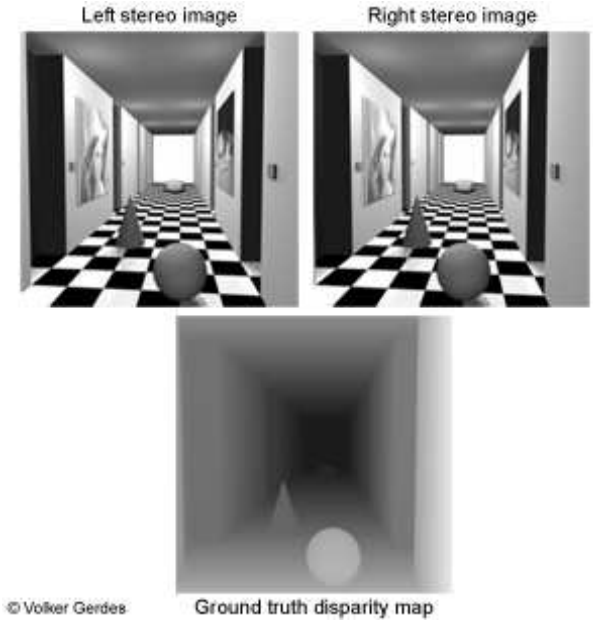

Figure 2.17: Stereo images and related disparity map

image from all the items such as pixels, edges, regions and objects in the left image [92, 93, 99]. The correspondence problem is commonly calculated with either area-based [100, 101, 24] or feature-based [92, 94, 102] methods. Feature based methods use features such as edges or points, like corners. Area based methods use neighbourhoods around the pixel being computed. The reconstruction problem needs additional information about the cameras and assumptions about the scene and uses them to estimate disparities between items [94, 103, 104, 105]. To reduce the searching time of finding corresponding items, one well-known constraint is the epipolar constraint [99] as illustrated in Fig. 2.18. A point in an image creates an epipolar line on which the corresponding point in the other image must lie. Fig. 2.18 shows an example of a 3D point P, the two camera locations Ol and Or and the point ml with its corresponding point mr. The line er and el are the epipolar lines and the plane formed by the points P, Ol and Or is the epipolar plane.

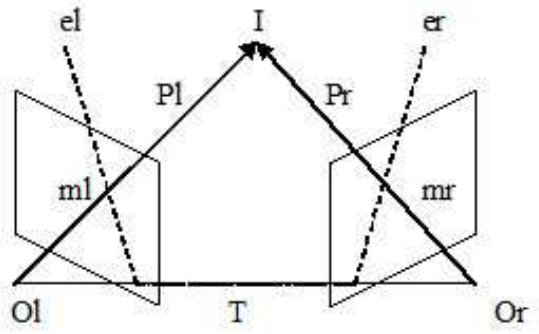

Figure 2.18: Epipolar geometry and epipolar plane

Epipolar geometry means a previous 2D search for corresponding points now becomes a 1D search, which as well as being much faster also means fewer false or ambiguous matches. This constraint to the correspondence search is called the epipolar constraint. In order to determine the epipolar line from a point in an image the essential matrix and the fundamental matrix must be calculated. The essential matrix defines the relationship between an image point defined in camera coordinates and the epipolar line. The fundamental matrix defines the relationship between an image point defined in pixel coordinates and the epipolar line. The essential matrix can be computed using three vectors that lie on the epipolar plane,P_l, T and $(P_l - T)$. The epipolar plane is computed by

$$(P_l - T)^T (T \times P_l) = 0 \rightarrow \left(R^T P_r\right)^T (T \times P_l) = 0 \tag{2.12}$$

Where $T \times P_l$ (cross product) is a vector perpendicular to the plane containing T and P_l (the epipolar plane). $P_l - T$ is also in the same plane, so the dot product is zero. The cross product can be written as

$$T \times P_l = SP_l \rightarrow S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \tag{2.13}$$

From 2.12 and 2.13 we get $P_r^T E P_l = 0 \rightarrow E = RS$ where matrix E is the essential matrix which defines the epipolar constraint in terms of the extrinsic parameters. More computation about epipolar constraints for stereo vision can be found in [99, 106]. As a basic constraint, many researchers have used the epipolar constraint for reducing search time [92]. Other constraints are also possible, for example the use of rigidity constraints [107].

Following the corresponding points process, depth information can be calculated, by using triangulation as illustrated in Fig.2.19. Triangulation is the process of determining the position of a point in a scene if that point is visible in two images [103]. Triangulation needs information about the camera for correct results, these are the focal length of the camera and the distance between the camera positions, or baseline, when the images were captured. These intrinsic camera properties can be found using camera calibration techniques. Suppose a point P is visible in two images. If the two camera matrices are known, the points ml and mr are projections of the point P in the two images. From the available data the two rays in space corresponding to the two image points may easily be computed [103]. The triangulation problem is to find the intersection of the two lines in space. For example to find the Z value of the point $P\,(X, Y, Z)$. The corresponding points in the images are of the form $ml\,(x1, y1)$ and its matching point $mr\,(x2, y2)$ where the point mr in the right image corresponds to the point ml in the left image. In this example the disparity is on the x axis only (for example if the corresponding points were on all the epipolar lines and the images were corrected). If the focal length of the camera is f, $y1 = y2$ and the distance between the cameras is b then the 3D point $P\,(X, Y, Z)$ can be computed by the equation 2.14. The intrinsic and extrinsic parameters of the cameras need to be calibrated or modelled [104, 105, 108].

$$Z = \frac{fb}{x2 - x1} \tag{2.14}$$

As illustrated in Fig. 2.19, 3D surface reconstruction needs to apply feature ex-

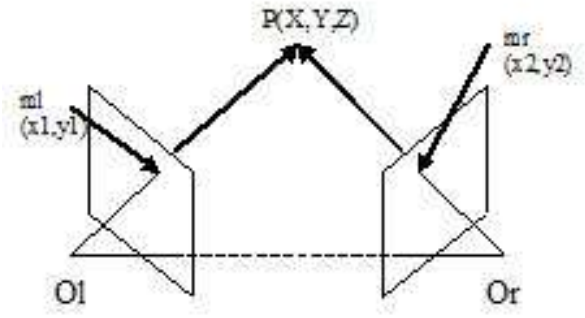

Figure 2.19: Triangulation to find a point in 3D from 2 2D images

traction for representing points, lines or areas for their correspondence search under different constraints. In addition to the epipolar line constraints, other constraints used by researchers include spatial continuity, disparity gradient and left-right consistency [100]. Spatial continuity attempts to ensure that the variation in depth over the image is smooth, with exceptions for discontinuities. The disparity gradient constraint limits the ratio of difference in disparity to difference in 3D position and can assist in finding false matches. For example the disparity gradient could be limited to a gradient of +1. The left-right consistency constraint states that feasible conjugate pairs are those found with both direct and reverse matching. For example for each point (x, y) on the left image the disparity $dl(x, y)$ is computed. The process is repeated with reversed images to compute $dr(x, y)$. If $dl\,(x, y) = -dr\,(x + dl\,(x, y)\,, y)$ then the point keeps its computed left disparity [99]. Examples of features used in stereo techniques are edges, silhouettes and interest points. Unlike silhouettes, interest points are very robust to changes in lighting, viewpoint, rotation and scaling which makes them very useful for the correspondence problem [49, 98, 109]. With knowledge of the epipolar geometry, the correspondence problem could be reduced from the original 2D search to a 1D search along the epipolar lines. However, the correspondence ambiguity with

the epipolar lines remains difficult to overcome, especially when the observed scene consists of dense features. Several factors make the stereo correspondence problem difficult: occlusions, large disparities, photometric and figural distortions, significant orientation difference of the image planes [37, 97, 110, 111, 49, 101] and the difficulty of deciding what window size to use in intensity-based matching [112, 113, 114, 83]. In addition to stereovision, a shape from shading algorithm [97] or motion is used to improve efficiency and accuracy of 3D. 3D reconstruction from stereovision is accurate but correspondence establishment is difficult. A complementary approach, structure-from-motion uses the motion cue. Motion analysis techniques use an image sequence, usually taken from a single video camera [95]. The camera is either moved around a static scene, the scene moves around a static camera, or in some cases both the camera and the scene move. Calculations are more difficult if both are moving. If objects within the scene are moving, for example moving traffic, then motion segmentation is required. The camera is usually moved slowly so that the distance between frames, the baseline, is small. This means the information in the image sequences moves little from one frame to the next. The motion of a camera relative to a static object can be described by a translation velocity vector and an angular velocity vector. There are two main methods for determining motion and structure from image sequences, optical flow and feature based methods. Structure from motion and structure from stereo have the same sub problems: the correspondence problem and the reconstruction problem. The motion cue has the advantage that the correspondence problem is relatively easy to solve because adjacent images are 'alike'. Thanks to video compression and media communication, algorithms for motion cue e.g. motion vectors for MPEG4 or H263 have been developed, standardised and implemented in hardware which makes the approach attractive and possible for real-time applications [96, 115, 72, 116, 117, 118]. The shortcomings of both structure-from-stereo and structure-from-motion techniques for 3D reconstruction have motivated a new direction of research to combine the two vision cues for 3D reconstruction when a stereo

pair of cameras are available to capture binocular image sequences [119]. Recently a few researchers have extended stereo imaging with panoramic imaging to create stereo panoramic imaging [120, 23, 25, 102, 121, 24], which lends itself to creating 3D environments of the real world.

2.5.3 Stereo Panoramas

A recent development in panoramic imaging is the stereo panorama. Stereo panoramas involve two panoramas, one for the left eye and one for the right eye. Peleg et al [120] mosaic images from a single rotating camera, and then generate a stereo panorama from the mosaic. Huang and Hung [23] base their panoramic stereo system on two cameras, which produce stereo panoramas for VR worlds. Pritch et al [120] presents a lens system and a mirror system for omni stereo panoramas. Peleg et al [120] also provides ideas about cameras that can be used for stereo panoramas. The next level is to use the stereo vision techniques and the stereo panoramic techniques to build 3D models from the panoramic data. Svoboda and Pajdla [122] reviewed current panoramic cameras that could be used for 3D computation. Koyasu et al [123] use omni directional cameras to build panoramas which are then compared to build 3D information and which is then used for robot navigation in dynamic environments. If panoramas are taken on the same horizontal plane problems with focus of contraction and focus of expansion arrise. If the panoramas are taken with vertical seperation, they can not be used for human stereo viewing but stereo imaging for depth calculation becomes much easier. Using panoramic cameras for 3D computation would enable full 3D scenes to be captured and modelled. There are limitations however with current systems. 3D models that are generated from stereo images or panoramas do not allow much movement, the view soon becomes distorted. To overcome this, multiple models must be built and combined to create a larger more complete model. Aliaga and Carlbom [90] have worked along these lines but no models are created, the system is based on image based rendering techniques.

2.5.4 Rendering and 3D visualisation

Once the panorama has been stitched to create the complete 360° image it needs to be visualised. There are a number of online viewers available for free download [11, 18]. The viewers remap the planar image onto the inside of a cylinder, sphere or cube. A cylinder is the simplest model because the planar image maps directly onto the inside. Because the panoramas need special viewers, they cannot take advantage of any hardware texture-mapping acceleration without approximating the cylinder's or sphere's shape with a polyhedron, which introduces distortions into the rendering. Szeliski and Shum [124] use environment maps to overcome this. Lee et al [125] present the idea of transmitting the panoramas in two levels, the background level, which is a static panorama, therefore using less information than a video panorama. The second level is a dynamic object overlaid on the background panorama. This enables pseudo video panoramas to be transmitted over a network, using less data than a pure video panorama. Another use of panoramic video is for tracking a region of interest, for example if a whole scene is captured but only the region of interest video stream is created from this to generate a standard video stream. Sun et al [126] use this system for tracking a speaker at a lecture and creating a single stream which pans with the movement of the speaker, similar to how a human operator would control the camera. For the stereo images that have generated a depth map, the depth map describes what shape the objects in the scene are, and can therefore build a 3D model of the scene. This scene can then be rendered as any 3D object would be. Texture information can be extracted from the individual images, and mapped to the 3D model, giving a realistic model of the scene. Another way to render the stereo information is using image based rendering [127]. Image-based rendering effectively deforms the image using homograph to give the effect of a 3D model. Shum and Kang [127] completed a review of current image-based rendering techniques with the conclusion that image-based rendering and more traditional model

based rendering have many complimentary characteristics, and in the future hardware should be designed for both. Kiyokawa et al [128] use an optical see-through display for mutual occlusion with a real-time stereovision system. This enables 3D objects to be integrated with video in real-time.

2.6 Discussions

Panoramic or omni-directional cameras allow the opportunity to capture rich data about an environment that can then be used to generate "walk-throughs" for VR, or backgrounds for games, or tourism [12]. The panoramic imaging system is a typical system using the convergence of computer vision and computer graphics as described above. Computer graphics and computer vision could be described as taking opposite approaches to the same problems. Computer vision develops novel capture techniques while computer graphics adopts techniques from computer vision for capturing models from the real world and also for reconstructing movement for virtual worlds. However, traditional view computer graphics start with inputting geometric models and producing image sequences whereas computer vision starts with inputting image sequences and producing geometric models, at least as an intermediary step. Linking with the real world, the virtual world is based on 3D, which will drive current panoramic imaging systems from 2D to 3D including hardware and software. The nucleation of virtual reality, graphics, video and vision is predicted to be an important area of research, particularly for 3D panoramic imaging and virtual environment construction. Some challenges exist in the following areas.

2.6.1 Stereo Vision Based Panoramic Capture System

It is well-known that stereo vision can obtain depth information for 3D scene capturing [79]. Most computer vision research to date has been concerned with the geometric recovery of points that can be matched between images [37]. Little work exists on

the problem of automatically recovering useful surface or image-based models from this data for stereo panoramic imaging. Instead of using special equipment, 3D scenes obtained from stereo vision can be easily visualised in a VR environment. Equally, the viewed scene can be changed based on different viewpoints and other view conditions. A very interesting and exciting application of mixing stereo and panoramic imaging is the Beagle 2 Stereo camera system [129]. Unfortunately it was never used on Mars, but they integrated stereo imaging for digital elevation models and parabolic mirrors for panoramic imaging into a single system. Video based reconstruction techniques must be developed which allow a user to interactively recover models of a scene and select viewpoints (much like a video paint brush that allows a user to interactively recover representations of the scene).

2.6.2 Deriving Capturing Conditions and Object Surface Characteristics

Image formation on a digital sensor integrates illumination information and object surface characteristics within the visible spectrum. It would be useful to reverse the capture device, illumination and scene surface characteristics from the captured data. For example, colour constancy algorithms have provided some approaches to estimate the illuminant of an image. Based on the illumination, linear models, particularly the diagonal model can be used for colour correction and transformation [130] where lighting or viewpoints are changed.

2.6.3 Image Understanding and 3D Reconstruction

Multiple view images for image sequences or different cameras have some correlated information. Based on the correlation, 3D images can be reconstructed. Further approaches may expand current registration techniques from correspondence points to correspondence area with the integration of shape, texture and colour. The representation of robust invariance will reduce the importance of accurate calibration on capturing systems. Augmented reality, as applied to the field of telerobotics, is

concerned with enabling a human operator to conduct tasks more effectively in a remote or hazardous environment when using a telepresence interface. Since vision is central to comprehending the remote environment, the main AR technique is to overlay computer generated graphical information upon the operator's view of the real scene [128]. Thus, it is possible to provide additional qualitative and quantitative information to the operator. In the latter case, the real and graphical worlds must be made to register, or correspond, with each other statically or dynamically depending on the application. Registration is required whenever quantitative information needs to flow between the real and graphical worlds and is a key element of most applications. It requires careful calibration and modelling of all the real world sensors into the graphical world. The sensor data is used to update the graphical model, often in real-time, using transformation matrices. Inaccuracies in the sensor data and the matrices gives rise to a dynamic registration error between the two worlds that manifests as a jerky or swimming motion for the overlaid graphics. Some invariant features for registration or correspondence, which are robust to image capturing condition and devices, will be a future research direction [130]. Video research is an evolution of vision and graphics work. The number of images is large and increasing bringing a need for greater compression. While it is good to have fewer samples, this does not guarantee fewer bits. If a sequence of images is seen as a video sequence, then general video coding can be applied, however, this intra-coding does not exploit the correlation between images. The use of intracoding however does not provide random access (i.e. frame N depends on frame N-1).

2.6.4 Image Based Volumetric Rendering

The convergence of computer vision, computer graphics and digital video technology has resulted in an emerging research area known as image-based rendering (IBR) [127]. In image-based modelling, rendering and animation, virtual environments and objects can be modelled and rendered directly from images and videos, bypassing

the difficult and labour-intensive processes of traditional model construction. Critical to image-based modelling and rendering of virtual environments, and image-based animation of virtual objects, are compression and decompression of the large amount of visual data. There is also no current work integrating virtual character animation with IBR, although on the horizon are hybrid representations that use both geometric models and IBR that will allow greater flexibility in dynamics. Also the integration of IBR with video needs to be addressed to add dynamic surface appearance. Volumetric reconstruction from multiviews is now quite well understood and analysed. The main difficulties remain in turning the volumetric representation into a model and/or IBR form and in extracting a representation of surface appearance properties to allow realistic rendering for 3D panoramas.

2.6.5 Problems identified

In this chapter the different stages of panoramic imaging and stereo vision techniques have been discussed. Major advances in digital imaging and vision computing have been reviewed. These reviews suggest direction for future research. 3D panoramic imaging will be a feasible approach for fast, realistic virtual environment construction. From the literature survey it was determined that the work should concentrate on several issues.

The first is the problem of colour constancy. When moving the camera to capture the next image in a panoramic sequence changes in the lighting and/or camera settings (automated) means that sequential images can have different appearances. Making sure the images are as similar as possible in colour and lighting is important for successful mosaicing.

The second problem is that of accurate image mosaicing. Determining an accurate transform matrix is essential for accurate mosaicing and the problem of correspondence is ongoing. The correspondence problem is an ongoing problem in the image processing research community, however for mosaicing only a small number of corre-

spondences are required and can be taken from areas where probability of a successful match will be high.

The third problem researched was that of accurate correspondence in areas of low texture. For example for 3D reconstruction of a surface with low texture there will be many correspondence errors.

Chapter 3

THEORY AND PROJECT ISSUES

3.1 Project concept and major elements

The basic theory of this project is to develop algorithms for areas within the process of creating 3D from 2D. The work concentrates on several small elements of the whole process. The project areas include the capture of the digital images, the pre-processing of the images, the 3D reconstruction and visualisation, as shown in Fig. 3.1.

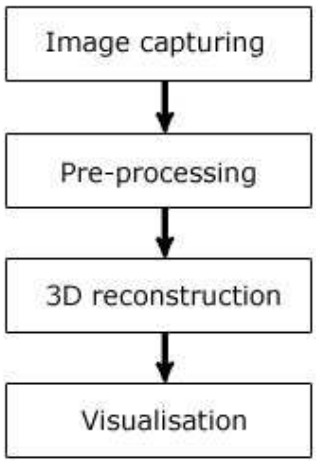

Figure 3.1: The basic theory of the project

Each of the four stages can be broken down into more detailed elements. The capture stage includes alignment of cameras, camera calibration and structured light pattern generation and projection. A human element exists at this stage which controls the focusing of the cameras and alignment. Templates can be used to overcome

the alignment problem and printed patterns used to achieve accurate focusing. The pre-processing includes epipolar alignment, colour correction, accurate mosaicing using homography and coordinate transform for the 'one-shot' mirror system. 3D reconstruction includes triangulation and accurate correspondence. For visualisation accurate texture recovery is required. Fig. 3.2 shows a more complete picture of the process, with each sub element shown under the main project stages. The shaded components are the areas this work addresses.

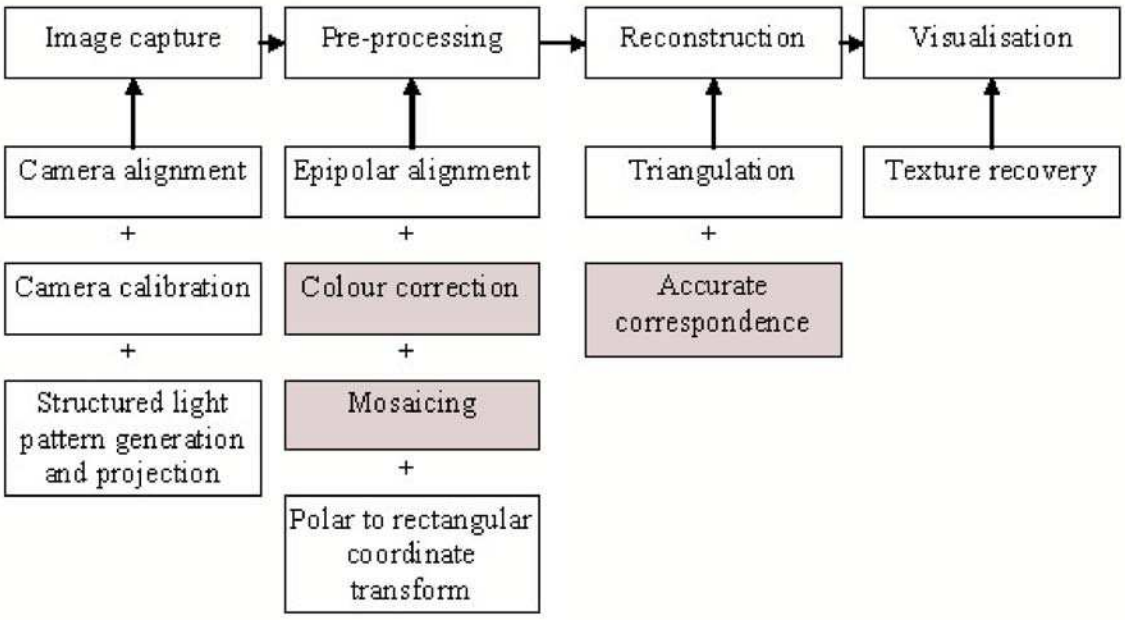

Figure 3.2: Showing the sub elements within each stage of the project

3.2 Key project elements

The key project elements are capture, processing, reconstruction and visualisation. Each one has key sub elements to be solved.

3.2.1 Image capture

Within the capture element accurate alignment of the cameras and camera calibration is required. Without accurate initial capture information the later stages will contain errors. It is important to ensure accurate capture by checking the alignment with templates, and camera calibration using a known calibration pattern, captured several times and used to calculate the camera properties. In the project two capture systems are available. One system is based on a high resolution digital SLR camera with a 0-360.com 'one-shot' parabolic mirror attached. This system enables the capture of panoramic images in a single image. The other system is a modular system. The modular system is able to capture stereo vision images, for example for object or face reconstruction, and also capture multiple images for mosaicing into a single panorama. This second system uses low resolution web cameras. The use of web cameras keeps the project costs to a minimum with the goal of proving accurate depth information can be achieved with low cost low resolution equipment. Camera calibration is only required for the second capture system. For example when reconstructing small objects using stereo vision, or capturing multiple images for mosaicing, camera calibration is required to determine the intrinsic and extrinsic parameters of the cameras. Another element of the capture phase is a structured light pattern. This pattern is used to assist the correspondence search algorithm by increasing the texture in areas of low texture, for example walls. The structured light pattern needs to be of sufficient size and distribution to assist the correspondence search, but also suitable for filtering for later texture recovery.

3.2.2 Pre-processing

In the pre-processing element the major sub elements are colour correction, mosaicing and epipolar alignment. The polar to rectangular conversion can be achieved using known algorithms either in commercial software, or by implementation in Mat-

lab. Colour correction is required for seamless mosaicing of the captured data, and also to assist the correspondence search, which uses pixel intensities to achieve correspondence. Colour correction for seamless mosaicing is required for visually accurate rendering of the panorama. Colour correction is also used before stereo vision processing to ensure similarity between the captured images. If the colour in the images is too dissimilar then errors in the correspondence search will occur. The mosaicing has to be accurate because the correspondence search will expect epipolar lines to be undistorted. Incorrect alignment of the images for mosaicing will create deformations in the final panoramic image, in which case the correspondence search will fail.

3.2.3 3D reconstruction

The key sub element in the reconstruction phase is the reconstruction algorithm. If the stereo system cameras are aligned in parallel then simple triangulation can be used. If the cameras are not aligned in parallel then a more complex algorithm is used. With the 'one-shot' system a depth calculation based on correspondence is used. In all reconstruction situations accurate correspondence is essential for accurate results. To aid the correspondence search constraints are used. The most common constraint is the epipolar constraint, which is used to limit the correspondence search to a 1D search instead of a 2D search.

3.2.4 Visualisation

For a visually pleasing and accurate visualisation of the final model a photo realistic texture must be used. The structured light pattern needs to be removed from the captured image data to produce the texture. To recover the texture, filtering techniques can be used to remove the structured light pattern. In a static scene it is also possible to capture a texture without the structured light pattern for a more accurate representation.

3.3 Methodology

The areas concentrated on in this work are that of colour correction in Chapter 5, mosaicing in Chapter 6 and correspondence search speed and accuracy in Chapter 7. The first problem tackled is that of colour correction. Current techniques, for example the diagonal model, are tested against linear and linear plus affine transform matrices to determine if they are more accurate at colour correction. Colour correction is important both for aesthetics and for correspondence searching. Mosaicing is also discussed, again for both aesthetics and correspondence searching for 3D reconstruction. Accurate correspondence is very important for an accurate homographic matrix for translating the images and work is developed using interest points. Methods for increasing the accuracy and speed of the correspondence search are also discussed. Structured light patterns are used for accuracy and a hybrid area and feature based constraint are used to increase speed.

Chapter 4

PANORAMIC STEREO IMAGING

4.1 Introduction

The previous chapters were a comprehensive survey in the areas of image processing, stereo imaging and panoramic imaging. The state-of-the-art panoramic image capturing and 3D virtual environment systems have been reviewed. Much progress has been made in the past few decades. However, more work is necessary to improve the efficiency and fidelity for building 3D environments to meet the requirements of potential users, for example entertainment such as video games, tourism and manufacturing such as collaborative design and validation. Based on the literature survey and industrial requirements, a new system is proposed for fast reconstruction of 3D environments using digital cameras or panoramic imaging capture systems. This chapter will discuss the system design and development and novel approaches in the system.

4.2 System Design

The proposed system is to be able to build image-based 3D environments by using stereo vision, including 'normal' digital cameras or a panoramic image capturing system. To reduce the system cost and improve the system efficiency and fidelity, advanced signal processing will be used for the image content understanding, image stitching and image based visualisation. The system framework is outlined in Fig. 4.1.

The system includes three components, which can be implemented by hardware and software. With the increase of computational power, the project includes design

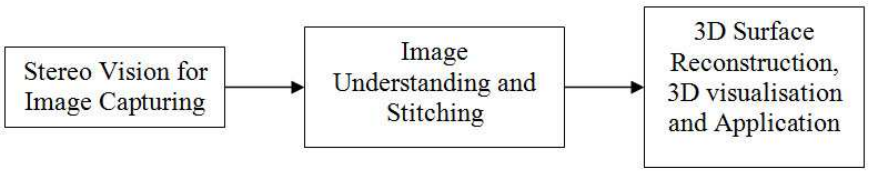

Figure 4.1: Stereovision system and panoramic imaging for 3D virtual environment

and development of image processing and handling algorithms for constructing virtual environments. The project includes the following:

- Evaluate current panoramic imaging systems and software
- Conceptualise a 3D panoramic imaging system
- Design and develop new image capturing system for 3D environments
- Design and implement a software environment for system investigation
- Design and develop new algorithms for image stitching
- Implement and evaluate robust feature-based correspondence points
- Investigate colour correction for images across viewpoints and devices
- Design and develop new algorithms for 3D surface reconstruction using stereo digital cameras or panoramic imaging devices
- Identify the characteristics and weaknesses of the developed system and software algorithms
- Propose further work based on the developed system and framework

The panoramic imaging systems in Chapter 2 are used by many people, commercially, academically and for fun. Many systems exist, each with its own advantages and disadvantages. Stereo imaging is also a popular area. The fusion of these two technologies though is still in its infancy. We propose new methods which could be used in a panoramic system whereby panoramic images are captured and processed to create 3D photorealistic environments suitable for user navigation. Fig. 4.2 shows a block diagram of the process for capturing and processing panoramic images to produce 3D information.

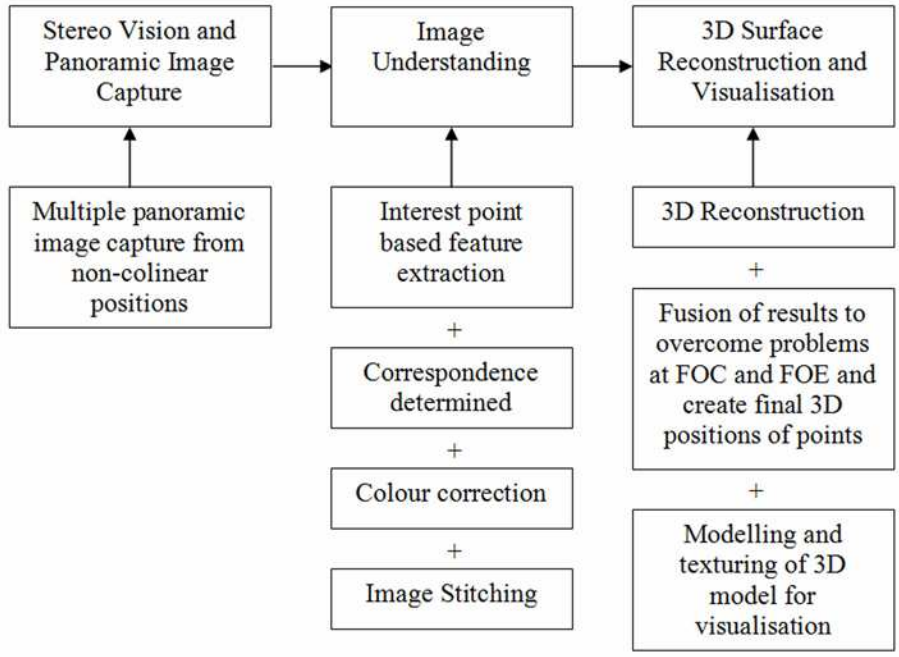

Figure 4.2: Showing the process of 3D panoramic imaging Where FOE is the Focus of Expansion and FOC is the Focus of Contraction, both parts of a panorama where either too much (FOE) or too little (FOC) depth is perceived.

The system used in the development of these new processes consists of a hardware element and a software element. The hardware element already exists in two forms, a one-shot system and a multi-shot system. Both systems can be used to generate

panoramic images, it is proposed 3D reconstruction processing can be achieved using those panoramic images. For the processing a software element is required. The initial processing (warping and stitching) is completed using commercial software, but later custom processing is developed in Matlab.

4.2.1 Hardware

There are two hardware systems, a 'normal' system and a panoramic system. The normal system consists of two off the shelf web cameras, these are used for the stereo processing and testing. This system is used for most of the testing, the stitching, colour correction and 3D surface reconstruction. The panoramic systems will be used in the future, although some testing will be done using these systems. The reason for first using the normal system is so that the principles behind the methods can be understood first, and then apply those to the panoramic imaging system later. The panoramic hardware used in the system takes two forms. The first hardware system captures a single cylindrical image. The second system captures 26 images, which are later stitched into a single spherical image. The reason for the two systems is to determine which produces the best results. Also the fact that the one-shot system only captures a cylindrical image means no depth information is available for the zenith and nadir. The first hardware system, which will be known as the 'one-shot' system uses a camera mounted vertically with a parabolic mirror mounted above it. The resulting image is spherical in appearance but is actually captured in polar coordinates, which are then mapped to rectilinear coordinates before processing. A picture of the one-shot system is shown in Fig. 4.3. The advantage of this system is that only one image is captured. Not only is this quicker than the second method, but it means dynamic scenes can be captured, for example moving traffic. The disadvantage is that the image is lower resolution using current image sensor technology and is limited in vertical field of view to approximately 115° . The second system, which will be known as the 'multi-shot' system uses a camera with a wide angle lens rotated about

its optical axis to cover all directions. 12 images are captured about the horizontal axis, then 6 images at +45° and 6 at -45° . 2 final images are captured, one pointing straight up (+90°) and one pointing down (-90°) for the zenith and nadir. The resulting images are stitched to create the finished sphere. A picture of the multi-shot system is shown in Fig. 4.4. The advantage of this system is the quality of the final panorama, it is approximately ten times the resolution of the one-shot system. Also more information is captured, including the ceiling and floor. Unfortunately dynamic scenes can not be captured. There is also a third system, used for testing the stereo correspondence algorithms, shown in Fig 4.5.

Figure 4.3: The one-shot system

Fig. 4.6 shows a diagram describing the camera orientations for the multi-shot system so that 26 images can be captured, ensuring everything is captured with sufficient overlap for accurate mosaicing. Blue arrows at horizontal, 12 images, 30° spacing. Red arrows 45° above horizontal, 6 images, 60° spacing. Yellow arrows 45° below horizontal, 6 images, 60° spacing. Black arrows, 1 image 90° above horizontal, 1 image

Figure 4.4: The multi-shot system, showing 4 of the 26 positions used

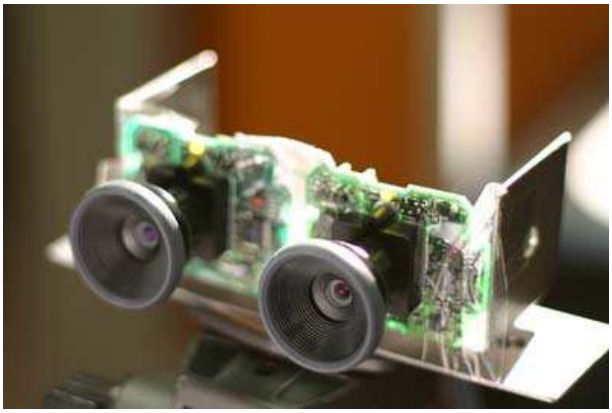

Figure 4.5: Stereo system using web cameras for stereo correspondence algorithm testing

90° below.

To capture panoramas suitable for 3D reconstruction processing, more than one panorama must be captured. Unless more than one camera system is available, dynamic scenes are not possible. The panoramas must also be captured in non-collinear positions. For example not all in a straight line. Non-collinear positioning of the camera systems is required because of the large field of view of the systems and to overcome the focus of contraction and the focus of expansion problems. When capturing panoramas using the 'one-shot' system the camera can be moved vertically. Because the zenith and nadir are not captured the focus of contraction and focus of expansion are not a problem.

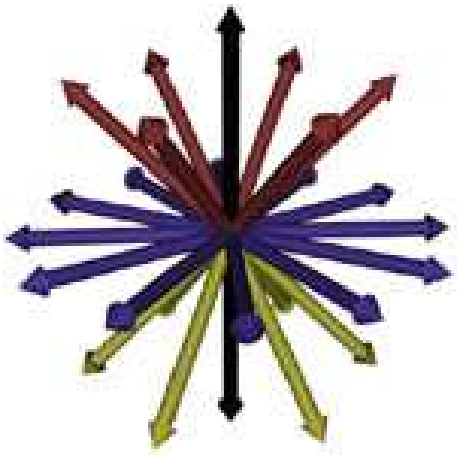

Figure 4.6: Diagram showing multi-shot system camera orientations.

4.2.2 Software

The proposed system also includes software development. For the un-warping and stitching of the captured images, commercially available software was initially used. Photoshop is used for the conversion from polar coordinates to rectilinear coordinates for the one-shot system. The multi-shot system images can be stitched with Panotools and Hugin/PTGui. All of the post-processing after the commercial software stages is completed using custom algorithms in Matlab. A graphical user interface (GUI) was also developed for ease of use and faster algorithm development and testing. The GUI includes abilities for importing images from the hardware system. The images are then processed using signal processing techniques. Processes such as colour correction, image stitching and 3D surface reconstruction are included. An interest point detector was the first algorithm developed. Interest points are used throughout the system, and therefore the most appropriate interest point detector is necessary. A modified Harris detector is used in this system. The modifications allow the choice of the number of interest points returned. Colour correction algorithms are also necessary. Part of the proposed GUI is able to detect and edit colour changes across panoramic

images using diagonal, diagonal plus affine, linear and linear plus affine transforms. The GUI is also able to process stereo information to produce 3D surfaces. The GUI includes feature based and area based methods. Fig 4.8 shows a diagram of the software framework for easy reference. Fig 4.7 shows an example of part of the software framework.

Figure 4.7: Screenshot of software framework.

4.3 Correspondence

As discussed in Chapter 2, correspondence is an important consideration for the project, indeed it is an ongoing problem in most computer vision research. With-

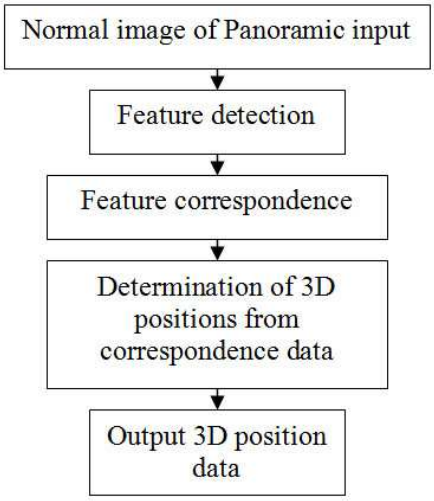

Figure 4.8: A diagram of the software framework

out the correct correspondence information it is impossible to obtain accurate 3D information. Many approaches have been proposed and used for the correspondence problem. The robustness and computing time are the main challenges for the problem. To develop a correspondence method robust to capture device, viewpoint and lighting variation, an interest point based method is proposed for the system. Correspondence between the panoramas is the most difficult and most important part of the 3D panoramic process. Correspondence is the linking of a point in one image with its corresponding point in the other image. There are two main methods of correspondence searching, area based methods and feature based methods. For this work a new interest point feature based correspondence method has been used. Interest points are points in the image where the pixel information changes two-dimensionally, for example at a corner, a t-junction, or a change in texture. Interest points can be detected by using equation 4.1:

$$\begin{aligned} \hat{C} &= \begin{vmatrix} \hat{I_x^2} & (\hat{I_x I_y}) \\ (\hat{I_x I_y}) & \hat{I_y^2} \end{vmatrix} \\ R(x,y) &= \det(\hat{C}) - k.trace^2(\hat{C}) \\ R(x,y) &> Threshold \Rightarrow Corner \end{aligned} \tag{4.1}$$

Where I_x and I_y are image intensity from horizontal and vertical direction Gaussian low pass filters.

Interest points will be used for image stitching. Their robustness to changes in lighting, viewpoint, scaling and rotation is well proven [45].

4.4 Content Based Colour Correction

One problem often encountered when capturing multiple images for stitching together is that of colour changes between images. Variables out of the users control such as lighting or capture device settings change. For example the two images shown in Fig. 4.9 are from consecutive images from a panoramic sequence and require, for aesthetic reasons, the same lighting results. The device settings have changed and therefore the colours have changed. A discussion of colour correction methods and the presentation of a new method are discussed in chapter 5. The colour correction methods discussed in chapter 5 are used in this system to combat any colour changes in the multi-shot system captured images.

4.5 3D Reconstruction

The 2D images that are captured using either the one- or multi-shot system do not have any depth information with the pixel data. The data stored includes colour information in a 2D array. The 2D array is filtered through a Bayer pattern and interpolated to form the RGB image. From this 2D information the system must be able to produce 3D results. This is achieved by using more than one 2D array of

Figure 4.9: Crops from consecutive images from a panorama sequence showing colour changes due to capture devices settings changes

image information. If a point in the scene can be acquired by more than one camera, and some information about the camera is known, then that point's 3D position can be established. The more images that are captured, the more accurate the results will be, and also points hidden to some cameras from some positions will be visible in other cameras, therefore giving a larger 3D representation. A discussion of a new 3D stereo imaging technique is discussed in Chapter 7. Chapter 7 also investigates efficiency of interest point based 3D surface reconstruction.

Chapter 5

COLOUR CORRECTION

Because interest points are robust to changes in viewpoint [45] they are used here to determine a linear transform to alter colours in consecutive images so they are more similar. We propose a method whereby the matching of histograms, independent of spatial information, is used to produce a transform which is then used to correct the colour information in another image. This chapter tests diagonal, diagonal plus affine, linear and linear plus affine transforms to determine the best possible result. We show that a linear model produces a better result.

5.1 Histogram Map Based Colour Correction

Colour and brightness variations often make it difficult to combine photographs into panoramic images. The individual source images and the overlapping regions remain distinguishable. This effect counteracts any effort to improve panorama resolution by using many photographs as sources. There are several origins for these variations even if the photographer takes care to keep constant exposure for each image: 1) Change of lighting conditions during the photo session; 2) Shutter speed variations; 3) Many more random and non reproducible parameters affecting image development and subsequent scanning. Even worse, many exposure related factors are outside the control of the photographer like automatic exposure settings in some cameras and scanners, or scanning services. To balance two stitching images, some colour correction is required. The normal approach applies the overlapping region of an image for estimating a colour transform matrix. Due to the variation of capturing conditions

such as camera resolution, noise and viewpoints, the pixels in the overlapping area are difficult to correspond to each other. In other words, the overlapping areas in two images may have different pixels due to variations of viewpoints and scaling. The difference of colour values at identical locations is not a very suitable optimization criterion. Real world images rarely fit together perfectly, and unavoidable spatial errors of one pixel or more between image 1 and image 2 may skew the optimization. To be independent of the spatial alignment of the images and apply general linear models (over 3 parameters in the transform matrix), it is intended that the matching of histograms of colours in the overlapping region of images 1 and 2 is used. Fig. 5.1 illustrates step 1 of this approach. A, B, C, D and A', B', C', D' are corresponding points in the two images I_1 and I_2, in which the maximum overlapping area is covered.

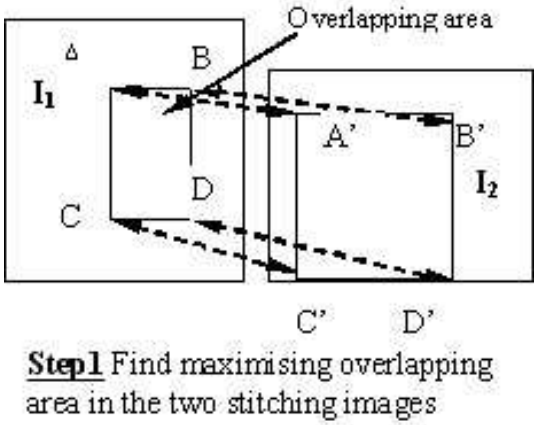

Figure 5.1: Step 1 of the new approach of colour correction

As illustrated in Fig. 5.1, the overlapped areas may have different histograms. We can map the histogram, where $A_1B_1C_1D_1$ area has the same 'geometrical content' as $ABCD$ in image I_1 but has the same histogram as $A'B'C'D'$ in image I_2.

In order to estimate M the images need to have the same corresponding pixels. In the case of panoramic imaging the viewpoint has changed and therefore the pixels

do not correspond. To overcome this, a correction is applied to the overlaps of the images. This correction creates two images with corresponding geometric information, but with different lighting information. The correction methods used are histogram equalisation, histogram mapping and singular value decomposition (SVD).

In Step 3, as shown in Fig. 5.3, we can apply the histogram mapped image for estimating colour transform matrix M, where $A_1B_1C_1D_1$ and $ABCD$ areas have the same image pixels with good correspondence. The transform matrix can be calculated by any linear transform as described in section 2.4. Finally, we can apply the transform matrix for the entire image I_1 or I_2 for colour calibration.

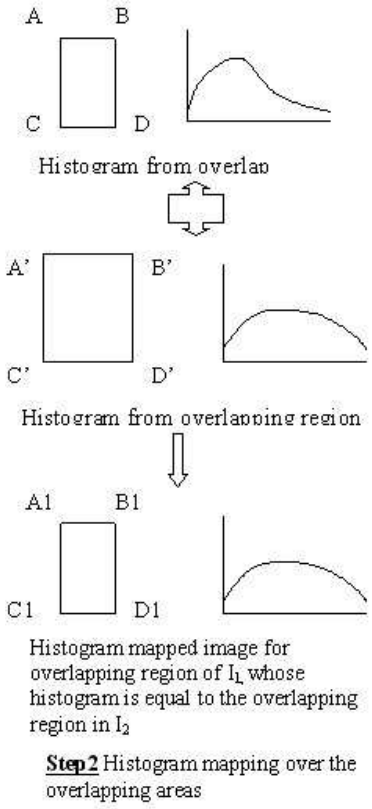

Figure 5.2: Step 2 of new approach of colour correction

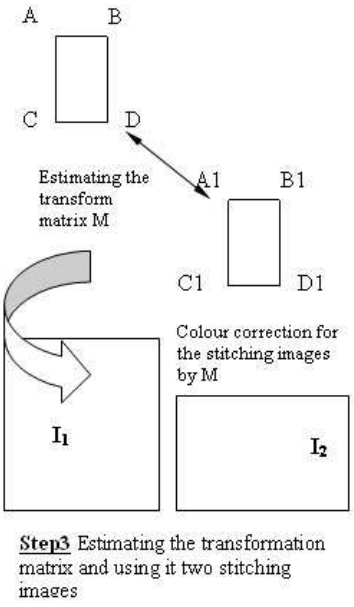

Figure 5.3: Linear transform matrix estimation and colour correction before image stitching and panorama

5.1.1 Histogram Equalisation

Histogram equalisation attempts to equalise the histogram as much as possible. Equalising or flattening a histogram often leads to a clearer image. For histogram equalisation in these tests the histograms of both images are equalised. If $r = 0$ represents black and $r = 1$ represents white and $0 \leq r \leq 1$ then for any r in the range 0 to 1, $s = T(r)$. From the previous equation a new pixel s will be produced for every pixel r in an image after a transform T has been applied. The probability density functions $P_r(r)$ and $P_s(s)$ characterise the original and transformed gray levels. If $P_r(r)$ and $T(r)$ are known and $T^{-1}(s)$ satisfies the condition that TI is single valued and monotonically increasing in the interval $0 \leq r \leq 1$, then the probability density function of the transformed gray levels is given by the relation

$$p_s(s) = \left[p_r(r) \frac{dr}{ds} \right]_{r=T^{-1}(s)} \tag{5.1}$$

Consider eq:5.2 where w is a dummy variable of integration. The result on the right side of eq:5.2 is known as the cumulative distribution function of r.

$$s = T(r) = \int_0^r p_r(w)\, dw \tag{5.2}$$

The conditions that TI is single valued and monotonically increasing in the interval $0 \leq r \leq 1$, and that $0 \leq T(r) \leq 1$ for $0 \leq r \leq 1$, are satisfied by the transformation function in eq:5.2 since the cumulative distribution function increases monotonically 0 to 1 as a function of r. From eq:5.2 the derivative of s with respect to r is given by

$$\frac{ds}{dr} = p_r(r) \tag{5.3}$$

If we then substitute dr/ds into eq:5.1 we get

$$p_s(s) = \left[p_r(r) \frac{1}{p_r(r)} \right]_{r=T^{-1}(s)} \tag{5.4}$$

$$p_s(s) = [1]_{r=T^{-1}(s)} \tag{5.5}$$

$=1\ 0 \leq r \leq 1$

which is a uniform density in the interval of definition of the transformed variable s. The above developments indicate that using a transformation function equal to the cumulative distribution of r produces and image whose gray levels have a uniform density. The transform function used to equalise the images is inverted for the second stage If M_1 and M_2 are the transformation functions of the images to equalise them, then $I = I_2 * M_1 * M_2^{-1}$ where I_2 is the uncorrected overlap from Image 2 and I is the new corrected I_2. An example of an equalised histogram is shown in Fig:5.4.

5.1.2 Histogram Mapping

Similar to histogram equalisation it is possible to specify a histogram that the image is desired to have. For this example we let $P_r(r)$ and $P_z(z)$ be the original and desired

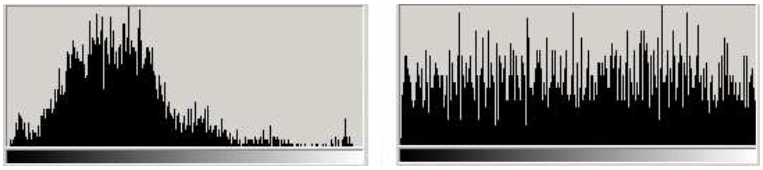

Figure 5.4: Histogram before and after equalisation

probability density functions respectively.

First imagine an image is histogram equalised:

$$s = T(r) = \int_0^r p_r(w)\, dw \tag{5.6}$$

And if we had the corrected image we could equalise it also:

$$s = G(z) = \int_0^z p_z(w)\, dw \tag{5.7}$$

and the inverse process, $z = G^{-1}(v)$ would give the desired levels back. Unfortunately the z levels are what we are trying to find, so the above equation is hypothetical. However, $P_s(s)$ and $P_v(v)$ would be identical uniform densities since the result of eq:5.2 is independent of the density inside the integral. So instead of using v in the inverse process we use s, the uniform levels obtained from the original image, the resulting levels, $z = G^{-1}(s)$, would have the desired probability function. To summarize the process:

1. Equalise the levels of the original image using eq:5.2

2. Specify the desired density function and obtain the transformation function $G(z)$ using eq:5.6

3. Apply the inverse transformation function, $z = G^{-1}(s)$, to the levels obtained in step 1

This procedure results in an image with the same geometrical content as the first image, but with the histogram specified from the second image.

5.1.3 Singular Value Decomposition

SVD, or singular value decomposition is an important matrix calculation. The SVD of an mxn matrix A gives two orthonormal matrices u and v and a diagonal matrix s such that: $A = usv^T$ u is an mxm matrix and v is a nxn matrix. s is rectangular with the same dimensions as A. The positive values of s are the singular values of A. The columns of u and v are called left and right singular vectors for A.

The pseudo code below goes through the process used to use SVD for correcting the colour in the images.

Image 1 $(I_1) = m * n$ pixels

Image 2 $(I_2) = m * n$ pixels

1) The image matrices (both are $[k, 3]$ matrices)

$$I_1 = \begin{bmatrix} R_1^1 & G_1^1 & B_1^1 \\ R_2^1 & G_2^1 & B_2^1 \\ \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow \\ R_k^1 & G_k^1 & B_k^1 \end{bmatrix} \text{ and } I_2 = \begin{bmatrix} R_1^2 & G_1^2 & B_1^2 \\ R_2^2 & G_2^2 & B_2^2 \\ \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow \\ R_k^2 & G_k^2 & B_k^2 \end{bmatrix}$$

Where $k = m * n$ and R^1, G^1, B^1 = Red, Green and Blue pixels from Image 1

2) $M_1 = I_1 - H * mean\,(I_1)$ and $N_1 = I_2 - H * mean\,(I_2)$

where $H = \begin{bmatrix} 1 \\ 1 \\ \downarrow \\ k \end{bmatrix}$

3) Calculate M_2 and N_2 [3,3] matrices

$M_2 = M_1^{'} * M_1$ and $N_2 = N_1^{'} * N_1$

4) $[u, s, v] = svd\,(M_2)$ and

$\left[u^N, s^N, v^N\right] = svd\left(N_2\right)$

5) Find maximum value and put in first row/column

$s\left(1,1\right) = \max\left(s\left(1,1\right), 5\right)$ and

$s^N\left(1,1\right) = \max\left(s^N\left(1,1\right), 5\right)$

6) $M_3 = pseudoinverse\left(\sqrt{s}\right) * u^{'}$ and

$N_3 = pseudoinverse\left(\sqrt{N}\right) * u^{N^{'}}$

7) $final = \left(pseudoinverse\left(M_3\right) * N_3 * N_1^{'}\right) + \left(mean\left(I_1\right) * \begin{bmatrix} 1 \\ 1 \\ \downarrow \\ k \end{bmatrix} \right)$

Find all values below 0 and make them equal 0, and similarly find all values over 255 and make them equal 255.

Reshape final into a [m,n,3] matrix and display as an image.

The result is image 2 now having the same lighting and colouring as image 1

5.2 Colour Correction for Panoramic Imaging

Based on the above-proposed approach, it has been applied for panoramic source image colour correction. Fig. 5.5 illustrates the overlapping area from two stitching images. A histogram map will be applied to the two overlapped images. Then, following the steps described in sections 2.4 and 5.1, transform matrices can be estimated by different linear models. Finally, apply the transform matrix for colour correction and image stitching. Fig. 5.6 shows the colour corrected images to image 1 or image 2 capture conditions. At the end of the chapter, Fig. 5.14 illustrates four panoramic images with the proposed colour correction approach. Linear transformation which considers the correlation of RGB channels can provide slightly better results. However, the linear model plus affine transform can provide excellent results.

Figure 5.5: Overlapping area of the stitching images

Figure 5.6: Colour correction based on M or $M-1$

5.3 Experimental Tests

Experiments to determine speed and accuracy of the different methods were conducted. A range of image domains were used, including both synthetic and real world. A variety of situations were simulated, including changes in brightness, changes in global lighting temperature and a mix of the two. A selection of the images used in the experiments can be seen in Fig. 5.7.

The colour difference was measured before and after each of the methods were applied, which is used to determine the accuracy of the methods. The colour difference is the average sum of squared differences for each colour channel.

Average (mean) for all image tests					
Average Initial Colour Difference = 0.39					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.57				
diagonal		0.006	0.013	0.204	40.41%
diagonal plus affine		0.082	0.054	0.190	44.30%
linear		0.020	0.025	0.198	42.37%
linear plus affine		0.028	0.112	0.190	44.34%
histogram equalisation	1.42				
diagonal		0.011	0.014	0.204	40.74%
diagonal plus affine		0.072	0.058	0.189	45.72%
linear		0.020	0.024	0.197	42.99%
linear plus affine		0.026	0.107	0.189	45.78%
svd	0.07				
diagonal		0.006	0.013	0.203	40.85%
diagonal plus affine		0.075	0.056	0.185	46.09%
linear		0.023	0.020	0.196	42.90%
linear plus affine		0.026	0.108	0.186	45.66%

Table 5.1: Average (mean) values for all colour correction image tests

Minimum values across all image tests					
Minimum Initial Colour Difference = 0.10					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.031				
diagonal		0.000	0.000	0.022	-0.72%
diagonal plus affine		0.000	0.000	0.044	3.14%
linear		0.000	0.000	0.030	1.32%
linear plus affine		0.000	0.078	0.044	3.18%
histogram equalisation	0.062				
diagonal		0.000	0.000	0.006	-1.18%
diagonal plus affine		0.016	0.031	0.007	1.28%
linear		0.000	0.000	0.013	0.56%
linear plus affine		0.000	0.078	0.007	1.31%
svd	0.000				
diagonal		0.000	0.000	0.006	-1.08%
diagonal plus affine		0.016	0.031	0.010	1.18%
linear		0.000	0.000	0.031	0.53%
linear plus affine		0.000	0.062	0.026	1.12%

Table 5.2: Minimum values for all colour correction image tests

Maximum values across all image tests					
Maximum Initial Colour Difference = 1.197					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	1.966				
diagonal		0.031	0.047	0.546	95.26%
diagonal plus affine		0.234	0.094	0.546	95.26%
linear		0.047	0.047	0.546	95.20%
linear plus affine		0.062	0.156	0.546	95.21%
histogram equalisation	5.335				
diagonal		0.047	0.031	0.566	96.90%
diagonal plus affine		0.187	0.094	0.602	96.50%
linear		0.047	0.031	0.593	95.28%
linear plus affine		0.078	0.140	0.602	96.45%
svd	0.250				
diagonal		0.031	0.031	0.547	96.86%
diagonal plus affine		0.187	0.094	0.538	95.79%
linear		0.078	0.047	0.538	95.23%
linear plus affine		0.062	0.140	0.538	95.61%

Table 5.3: Maximum values for all colour correction image tests

Standard Deviation across all image tests					
Standard Deviation in Colour Difference = 0.391					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.565				
diagonal		0.006	0.013	0.204	40.41%
diagonal plus affine		0.082	0.054	0.190	44.30%
linear		0.020	0.025	0.198	42.37%
linear plus affine		0.027	0.112	0.190	44.34%
histogram equalisation	1.420				
diagonal		0.011	0.014	0.204	40.74%
diagonal plus affine		0.072	0.058	0.189	45.72%
linear		0.020	0.024	0.197	42.99%
linear plus affine		0.026	0.107	0.189	45.78%
svd	0.071				
diagonal		0.006	0.013	0.203	40.85%
diagonal plus affine		0.075	0.056	0.185	46.09%
linear		0.023	0.020	0.196	42.90%
linear plus affine		0.026	0.108	0.186	45.66%

Table 5.4: Standard Deviation for all colour correction image tests

Average (mean) values across the brightness variation test images					
Average Initial Colour Difference = 0.455					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.468				
diagonal		0.005	0.012	0.206	45.95%
diagonal plus affine		0.069	0.050	0.186	50.66%
linear		0.018	0.027	0.198	47.72%
linear plus affine		0.028	0.108	0.186	50.73%
histogram equalisation	1.206				
diagonal		0.009	0.015	0.206	46.02%
diagonal plus affine		0.067	0.058	0.187	50.89%
linear		0.016	0.026	0.198	48.27%
linear plus affine		0.023	0.107	0.187	50.96%
svd	0.062				
diagonal		0.003	0.014	0.205	46.13%
diagonal plus affine		0.068	0.054	0.183	51.29%
linear		0.019	0.021	0.194	48.57%
linear plus affine		0.021	0.107	0.183	51.42%

Table 5.5: Average (mean) values across the brightness variation colour correction test images

Minimum values across the brightness variation test images					
Minimum Initial Colour Difference = 0.150					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.031				
diagonal		0.000	0.000	0.048	1.97 %
diagonal plus affine		0.000	0.000	0.048	7.27 %
linear		0.000	0.016	0.049	1.43 %
linear plus affine		0.016	0.078	0.049	7.27 %
histogram equalisation	0.062				
diagonal		0.000	0.000	0.048	1.20 %
diagonal plus affine		0.016	0.031	0.039	2.33 %
linear		0.000	0.000	0.048	1.24 %
linear plus affine		0.000	0.078	0.039	2.36 %
svd	0.000				
diagonal		0.000	0.000	0.048	1.19 %
diagonal plus affine		0.016	0.031	0.043	2.14 %
linear		0.000	0.000	0.048	1.77 %
linear plus affine		0.000	0.062	0.045	2.58 %

Table 5.6: Minimum values across the brightness variation colour correction test images

Maximum values across the brightness variation test images					
Maximum Initial Colour Difference = 1.197					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.998				
diagonal		0.031	0.016	0.546	95.26 %
diagonal plus affine		0.156	0.094	0.546	95.26 %
linear		0.047	0.047	0.546	95.20 %
linear plus affine		0.047	0.140	0.546	95.21 %
histogram equalisation	2.777				
diagonal		0.031	0.031	0.566	95.25 %
diagonal plus affine		0.156	0.094	0.602	96.15 %
linear		0.031	0.031	0.593	95.28 %
linear plus affine		0.047	0.125	0.602	96.13 %
svd	0.125				
diagonal		0.016	0.016	0.547	95.27 %
diagonal plus affine		0.156	0.094	0.538	95.79 %
linear		0.047	0.031	0.538	95.23 %
linear plus affine		0.047	0.140	0.538	95.61 %

Table 5.7: Maximum values across the brightness variation colour correction test images

Standard Deviation values across the brightness variation test images					
Standard Deviation of Initial Colour Difference = 0.315					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.319				
diagonal		0.009	0.007	0.134	25.38 %
diagonal plus affine		0.037	0.024	0.123	24.06 %
linear		0.010	0.009	0.131	25.25 %
linear plus affine		0.013	0.015	0.123	24.09 %
histogram equalisation	0.901				
diagonal		0.011	0.006	0.138	25.60 %
diagonal plus affine		0.034	0.021	0.140	26.19 %
linear		0.011	0.009	0.140	25.98 %
linear plus affine		0.011	0.014	0.141	26.20 %
svd	0.032				
diagonal		0.006	0.005	0.135	25.64 %
diagonal plus affine		0.035	0.022	0.129	26.39 %
linear		0.011	0.012	0.131	26.07 %
linear plus affine		0.015	0.020	0.129	26.41 %

Table 5.8: Standard Deviation of values across the brightness variation colour correction test images

Average (mean) values across the white balance variation test images					
Average Initial Colour Difference = 0.270					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.682				
diagonal		0.008	0.016	0.190	28.90%
diagonal plus affine		0.097	0.059	0.178	34.01%
linear		0.023	0.022	0.182	32.39%
linear plus affine		0.031	0.114	0.178	34.03%
histogram equalisation	1.831				
diagonal		0.012	0.014	0.188	29.49%
diagonal plus affine		0.083	0.061	0.172	37.39%
linear		0.019	0.023	0.180	33.45%
linear plus affine		0.031	0.108	0.172	37.48%
svd	0.076				
diagonal		0.006	0.014	0.188	29.62%
diagonal plus affine		0.094	0.059	0.170	37.89%
linear		0.022	0.020	0.181	33.46%
linear plus affine		0.031	0.106	0.174	36.98%

Table 5.9: Average (mean) values across the white balance variation colour correction test images

Minimum values across the white balance variation test images					
Minimum Initial Colour Difference = 0.101					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.094				
diagonal		0.000	0.000	0.058	-0.72 %
diagonal plus affine		0.016	0.031	0.044	3.14 %
linear		0.000	0.000	0.047	1.32 %
linear plus affine		0.016	0.094	0.044	3.18 %
histogram equalisation	0.218				
diagonal		0.000	0.000	0.058	-1.18 %
diagonal plus affine		0.016	0.031	0.020	1.28 %
linear		0.000	0.016	0.039	0.56 %
linear plus affine		0.016	0.094	0.020	1.31 %
svd	0.031				
diagonal		0.000	0.000	0.058	-1.08 %
diagonal plus affine		0.047	0.031	0.020	1.18 %
linear		0.000	0.000	0.039	0.53 %
linear plus affine		0.000	0.078	0.026	1.12 %

Table 5.10: Minimum values across the white balance variation colour correction test images

Maximum values across the white balance variation test images					
Maximum Initial Colour Difference = 0.686					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	1.966				
diagonal		0.031	0.047	0.452	73.45 %
diagonal plus affine		0.234	0.094	0.434	78.82 %
linear		0.047	0.031	0.434	71.41 %
linear plus affine		0.062	0.140	0.432	78.41 %
histogram equalisation	5.335				
diagonal		0.047	0.016	0.452	74.52 %
diagonal plus affine		0.187	0.094	0.425	82.34 %
linear		0.031	0.031	0.425	73.75 %
linear plus affine		0.078	0.140	0.425	82.12 %
svd	0.203				
diagonal		0.031	0.031	0.451	74.47 %
diagonal plus affine		0.187	0.094	0.415	82.29 %
linear		0.047	0.047	0.454	75.49 %
linear plus affine		0.062	0.140	0.456	84.32 %

Table 5.11: Maximum values across the white balance variation colour correction test images

Standard Deviation values across the white balance variation test images					
Standard Deviation of Initial Colour Difference = 0.155					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.519				
diagonal		0.011	0.015	0.110	19.23 %
diagonal plus affine		0.062	0.023	0.109	20.59 %
linear		0.017	0.013	0.106	19.17 %
linear plus affine		0.013	0.015	0.108	20.47 %
histogram equalisation	1.441				
diagonal		0.014	0.005	0.110	19.42 %
diagonal plus affine		0.048	0.021	0.111	25.24 %
linear		0.012	0.008	0.106	20.85 %
linear plus affine		0.019	0.016	0.111	25.22 %
svd	0.047				
diagonal		0.011	0.009	0.110	19.36 %
diagonal plus affine		0.045	0.024	0.108	25.23 %
linear		0.013	0.015	0.114	21.17 %
linear plus affine		0.018	0.018	0.119	24.61 %

Table 5.12: Standard Deviation of values across the white balance variation colour correction test images

Average (mean) values across the white balance and brightness variation test images					
Average Initial Colour Difference = 0.382					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.643				
diagonal		0.006	0.012	0.216	40.85 %
diagonal plus affine		0.094	0.056	0.210	41.85 %
linear		0.022	0.025	0.214	41.64 %
linear plus affine		0.022	0.117	0.210	41.87 %
histogram equalisation	1.438				
diagonal		0.014	0.011	0.216	41.45 %
diagonal plus affine		0.072	0.056	0.209	43.69 %
linear		0.027	0.020	0.215	41.95 %
linear plus affine		0.028	0.106	0.208	43.71 %
svd	0.086				
diagonal		0.012	0.011	0.215	41.51 %
diagonal plus affine		0.072	0.058	0.203	43.89 %
linear		0.031	0.017	0.215	41.01 %
linear plus affine		0.030	0.112	0.205	42.80 %

Table 5.13: Average (mean) values across the white balance and brightness variation colour correction test images

Minimum values across the white balance and brightness variation test images					
Minimum Initial Colour Difference = 0.151					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.203				
diagonal		0.000	0.000	0.022	1.92 %
diagonal plus affine		0.047	0.031	0.064	5.99 %
linear		0.016	0.016	0.030	1.57 %
linear plus affine		0.000	0.094	0.064	6.03 %
histogram equalisation	0.374				
diagonal		0.000	0.000	0.006	1.26 %
diagonal plus affine		0.016	0.031	0.007	2.51 %
linear		0.016	0.000	0.013	1.38 %
linear plus affine		0.016	0.094	0.007	2.51 %
svd	0.016				
diagonal		0.000	0.000	0.006	1.24 %
diagonal plus affine		0.031	0.031	0.010	2.37 %
linear		0.016	0.000	0.031	1.55 %
linear plus affine		0.016	0.094	0.033	2.12 %

Table 5.14: Minimum values across the white balance and brightness variation colour correction test images

Maximum values across the white balance and brightness variation test images					
Maximum Initial Colour Difference = 1.179					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	1.732				
diagonal		0.031	0.016	0.521	88.93 %
diagonal plus affine		0.203	0.094	0.521	68.19 %
linear		0.047	0.047	0.521	85.34 %
linear plus affine		0.062	0.156	0.521	68.19 %
histogram equalisation	3.853				
diagonal		0.031	0.031	0.534	96.90 %
diagonal plus affine		0.140	0.094	0.532	96.50 %
linear		0.047	0.031	0.538	93.80 %
linear plus affine		0.062	0.140	0.529	96.45 %
svd	0.250				
diagonal		0.031	0.016	0.519	96.86 %
diagonal plus affine		0.187	0.078	0.467	95.30 %
linear		0.078	0.031	0.518	84.48 %
linear plus affine		0.062	0.140	0.465	83.87 %

Table 5.15: Maximum values across the white balance and brightness variation colour correction test images

Standard Deviation values across the white balance and brightness variation test images					
Standard Deviation of Initial Colour Difference = 0.297					
	Time(s)	Calculate(s)	Apply(s)	Colour Difference	% Change
histogram mapping	0.480				
diagonal		0.011	0.007	0.141	23.62 %
diagonal plus affine		0.054	0.021	0.135	19.11 %
linear		0.011	0.011	0.141	22.82 %
linear plus affine		0.017	0.018	0.135	19.11 %
histogram equalisation	1.104				
diagonal		0.012	0.011	0.147	25.55 %
diagonal plus affine		0.041	0.023	0.147	25.22 %
linear		0.011	0.011	0.147	24.75 %
linear plus affine		0.016	0.018	0.146	25.22 %
svd	0.066				
diagonal		0.012	0.008	0.144	25.63 %
diagonal plus affine		0.046	0.020	0.131	25.42 %
linear		0.022	0.012	0.140	22.80 %
linear plus affine		0.017	0.016	0.128	23.05 %

Table 5.16: Standard Deviation values across the white balance and brightness variation colour correction test images

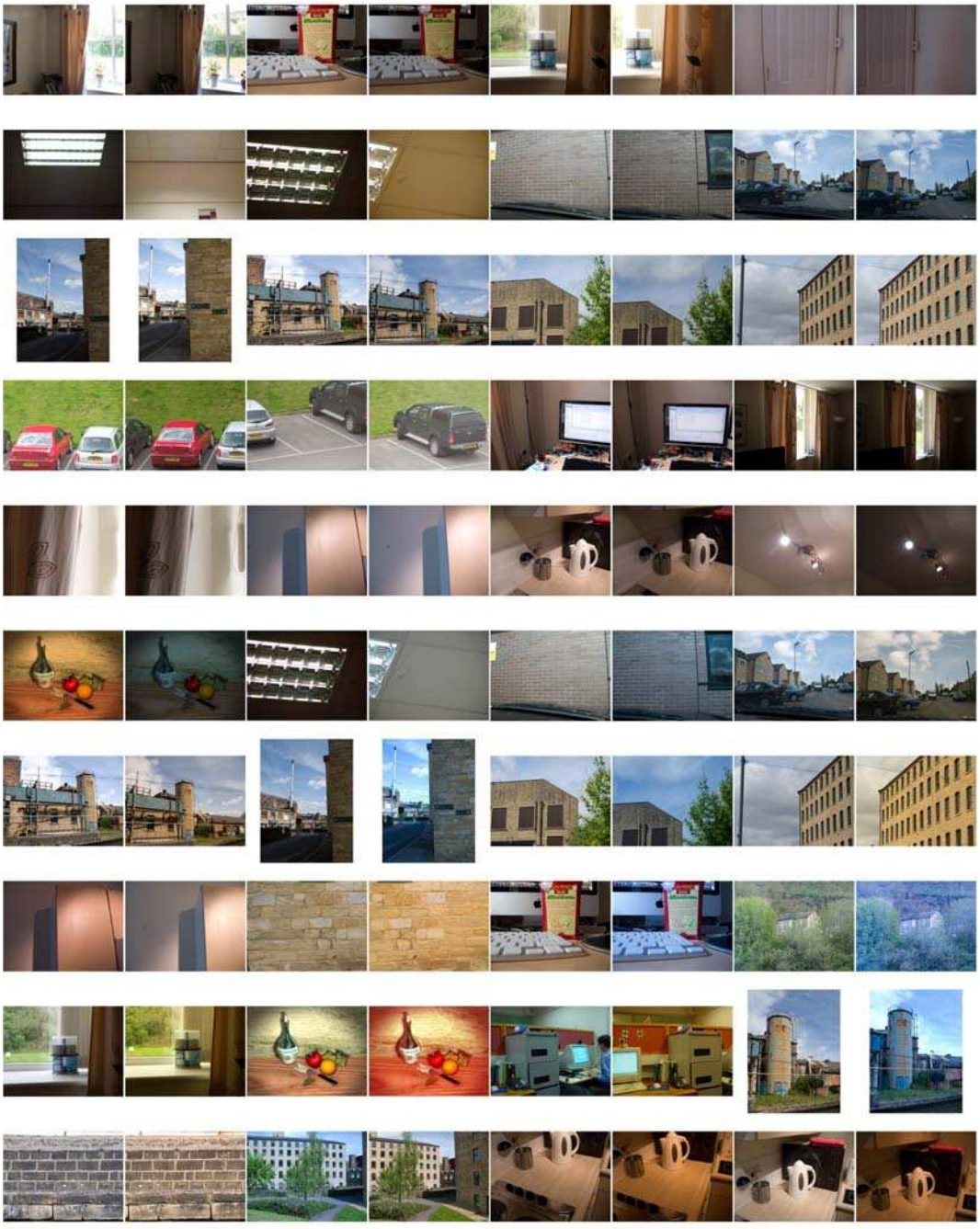

Figure 5.7: A selection of images used in the experimental tests for the colour correction algorithms

Results conclusions

As shown in Fig. 5.8, the fastest pre-processing method is SVD with an average of 0.07 seconds, followed by histogram mapping averaging 0.6 seconds and histogram equalisation averaging 1.5 seconds. The SVD approach might lend itself to real time situations averaging 14 frames per second.

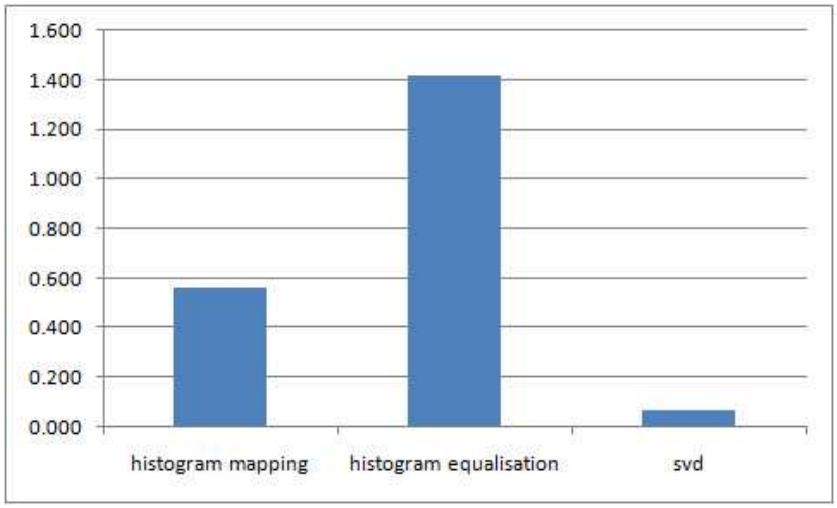

Figure 5.8: Comparison of the speed for the colour correction pre-processing methods

The affine correction proved to be advantageous over non-affine in the tests, as can be seen in Fig. 5.9. When affine correction is added to the diagonal model the accuracy of the results is significantly increased. The linear model proved more accurate than diagonal alone, but both were improved with the addition of affine. Using SVD as the pre-processing method also generally increased the accuracy.

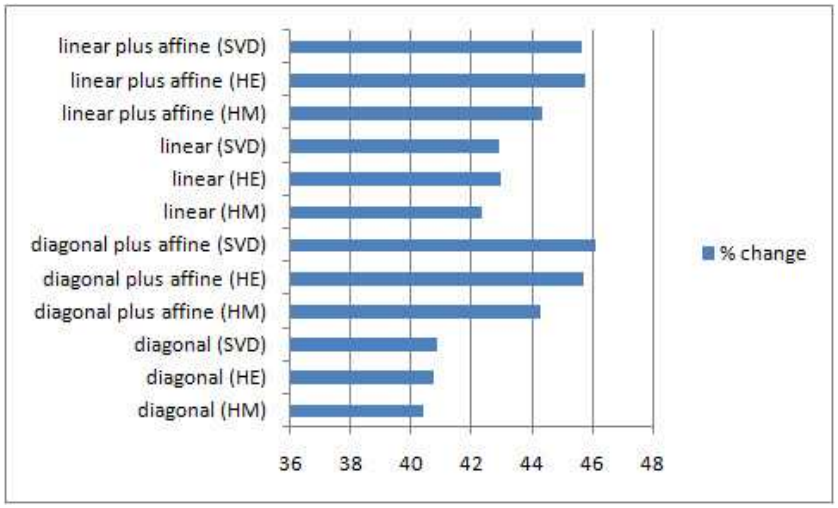

Figure 5.9: Comparison of accuracy for the colour correction processing methods

Fig. 5.10 shows the average speeds for each of the four methods. Adding affine to the calculations slowed the methods considerably. Although adding the affine correction increases accuracy of the methods, a system designer would need to weigh up whether that increase in accuracy was worth the associated increase in processing time, particularly important in a real time system.

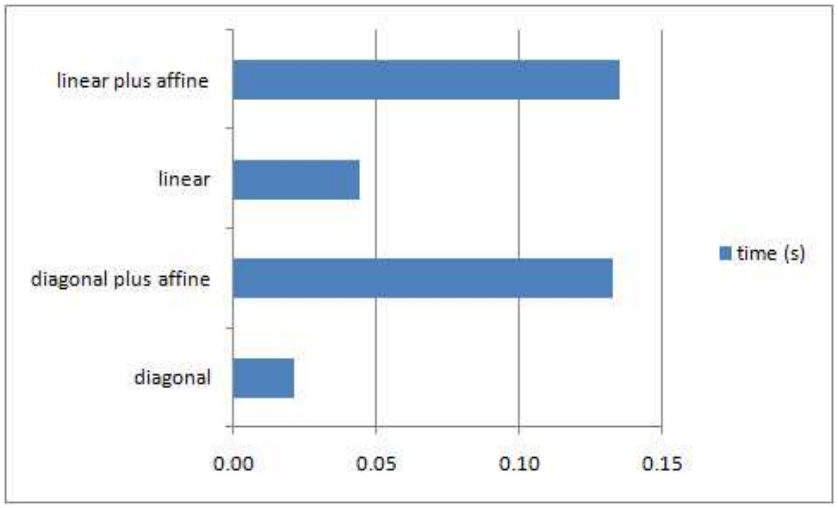

Figure 5.10: Comparison of the average speed of the colour correction methods

Fig. 5.11 shows the difference between the different groups of tests against the average for all tests. White balance changes proved more difficult to correct than brightness variations.

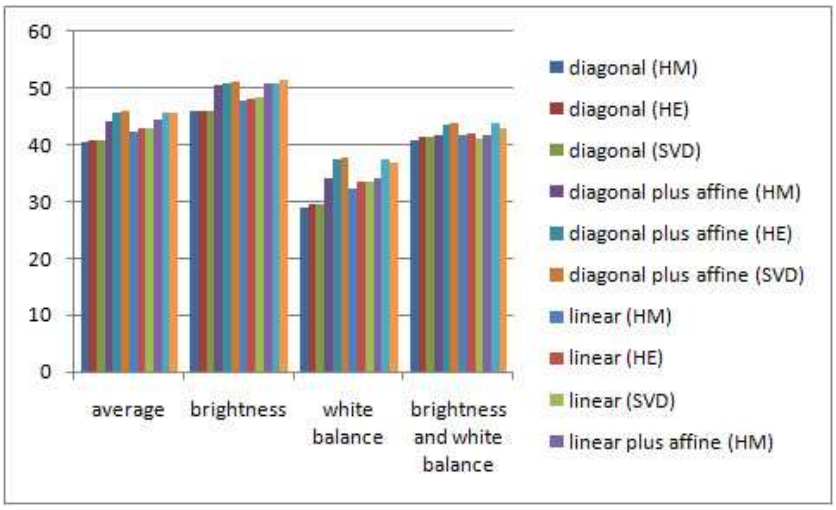

Figure 5.11: Comparison of the differences in processing accuracy for different image variations

Fig. 5.12 shows the stability of the pre-processing methods. The standard deviation for the SVD method is very low, suggesting a consistent processing time. Histogram equalisation is the least stable. This variation is further illustrated in Fig. 5.13.

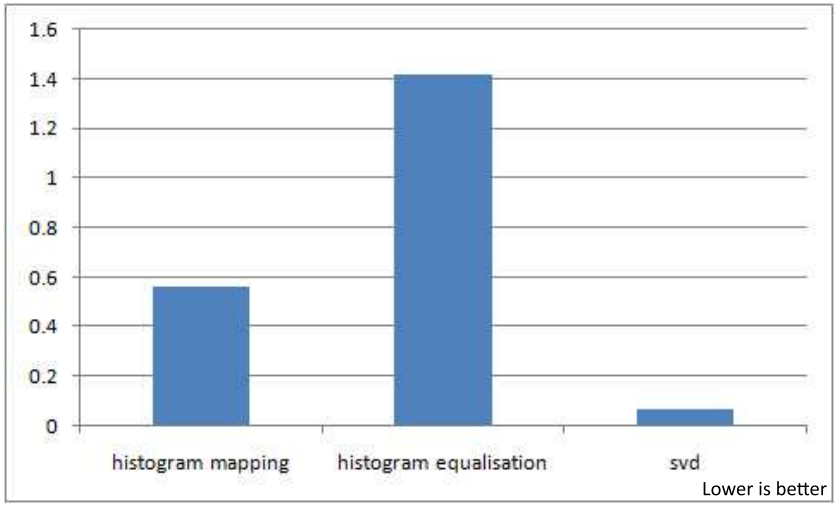

Figure 5.12: Comparison of the standard deviation in processing time for each pre-processing method

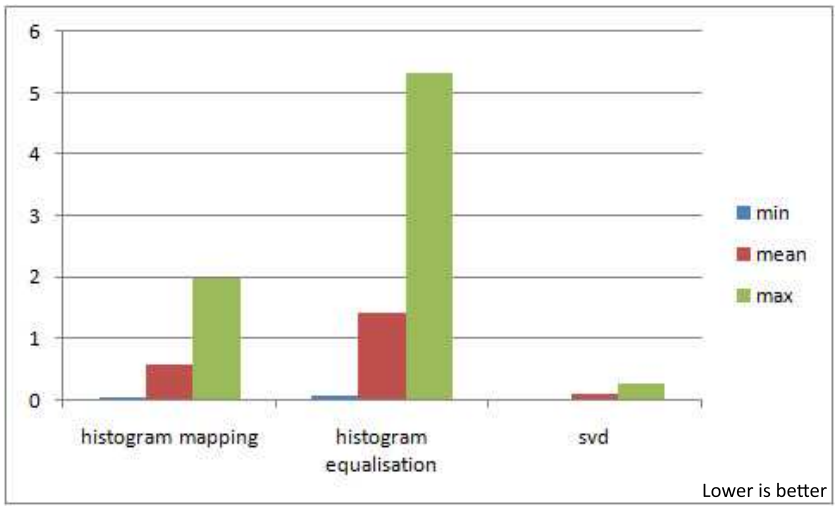

Figure 5.13: Comparison of the minimum, mean and maximum processing times for each pre-processing method

5.4 Summary

Colour correction for image mosaic based panoramic imaging is important. Several colour linear transformations have been discussed. The performance of linear models for colour correction is slightly better than the diagonal model, however the diagonal model is faster to calculate. The addition of affine to the models significantly increases the accuracy of the results, but with an increased processing time penalty. A histogram map based colour correction has been proposed and tested. Different colour correction based diagonal and linear transformations can be used in the proposed approach. The choice of linear or diagonal transformation matrix for colour correction can be affected by the accuracy and computing time. The results showed the SVD is the fastest method and that diagonal plus affine and linear plus affine are the most accurate colour correction methods.

a) Mini panorama without colour correction processing

b) Mini panorama using diagonal model plus affine colour correction

c) Mini panorama with linear correction pre-processing

d) Mini panorama with linear plus affine correction pre-processing

Figure 5.14: Different colour correction for panoramic imaging

Chapter 6

IMAGE MOSAICING

6.1 Introduction

An integral and important part of a 2D or 3D panoramic system is accurate image stitching or mosaicing. For normal devices image stitching is a necessary requirement, but the one-shot system requires no stitching. The one-shot is a mirror system which captures the whole scene in a single image, hence no stitching required. The resulting image from the mirror does require remapping from polar coordinates to rectilinear coordinates. Inaccurate stitching for multi shot systems would lead to inaccurate correspondence matching due to misalignment and therefore lead to inaccurate 3D modelling. In this chapter interest points are used to determine correspondence between consecutive images to produce an accurate stitch. A new 4-step interest point based method for image stitching is introduced, including how to identify the interest points, finding correspondence and then the spectral and spatial transforms used to position the images. Smoothing is also discussed.

6.2 4-step Interest Points Based Image Mosaic

The present study proposes a new methodology to map the geometric and spectral distribution of stitching images (Fig. 6.1). This work proposes a new 4-step model for construction of panoramic images. First it is necessary to identify the interest points in the image. Interest points are used because of their robustness to changes in viewpoint, lighting and scale [45]. Corresponding points in the images then have to be used for the transform matrices and at least four corresponding points will be

required. The intensity information at each of the interest points used to form the transform matrix will then be used to alter the image intensities to match each other. The details of the process are described below.

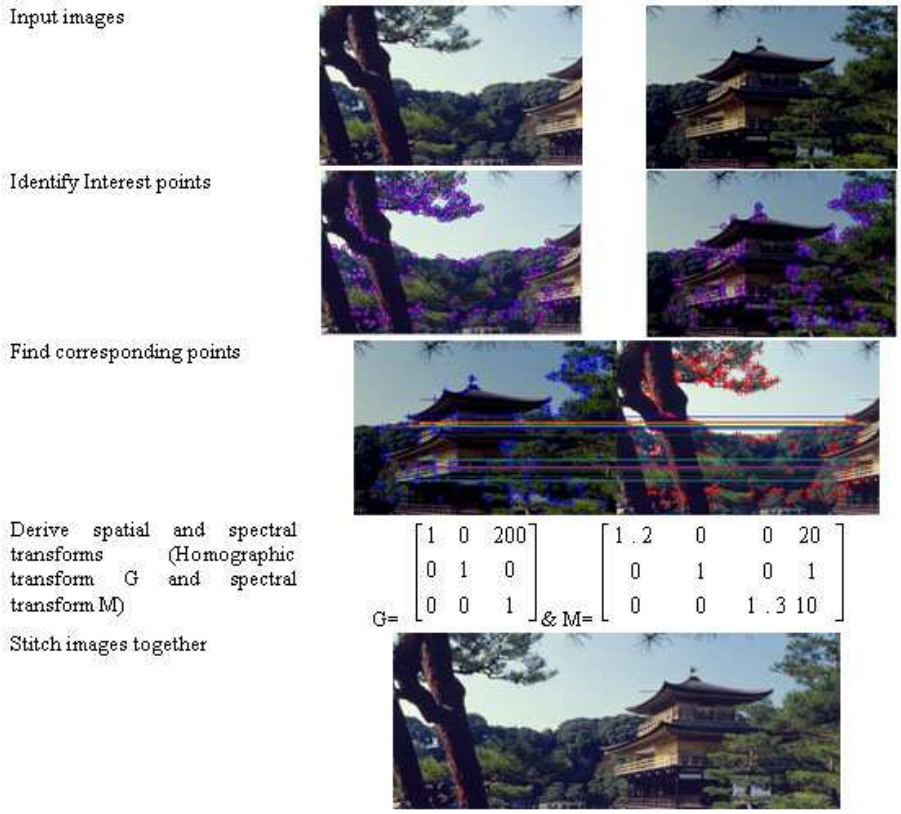

Figure 6.1: Image mosaic processing block-diagram

6.2.1 Identification of Interest Points

Interest points are used to represent image characteristics or content changes. There are two important requirements for feature points or interest points. Firstly, points corresponding to the same scene point should be extracted consistently over the different views, different lighting conditions and even different devices. If this were not the case, it would be impossible to find correspondences. Secondly, there should be

enough information in the neighbourhood of the points, so that corresponding points can be automatically matched. Many feature point extractors have been proposed.

In this system, an evaluation of the Harris corner detector is used. Consider the following matrix

$$M = \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \tag{6.1}$$

Where $I(x, y)$ is the grey level intensity. If at a certain point the two eigenvalues of the matrix are large, then a small motion in any direction will cause an important change of grey level. This indicates that the point is a corner. The corner response function is given by:

$$R = \det M - k(traceM)^2 \tag{6.2}$$

Where k is a parameter set to 0.04 (a suggestion of Harris). Corners are defined as local maxima of the "cornerness" function. Sub-pixel precision is achieved through a quadratic approximation in the neighbourhood of the local maxima. To avoid corners due to image noise, it can be interesting to smooth the images with a Gaussian filter using $\sigma = 0.6$. This should however not be done on the input images, but on images containing the squared image derivatives i.e.

$$\left(\frac{\partial I}{\partial x}\right)^2, \left(\frac{\partial I}{\partial y}\right)^2, \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \tag{6.3}$$

In practice far too many corners are often extracted. In this case it is often useful to first restrict the numbers of corners before trying to match them. For a colour image, the system processes the three RGB channels independently as grey scale images [56]. After the above evaluation for Harris corner detection, the detected interest points are robust to viewpoint change, lighting differences, rotation and scaling due to the process of various grey level derivatives and using local maxima as response function

in eq. 6.2. Although RGB colour space is less robust to lighting differences than hue or saturation components, the interest points are robust to light variation. Fig. 6.2 illustrates the interest points on different images produced when the same object was captured under spectral illumination with different points, where the colour images are from a colour constancy database [44]. As illustrated in Fig. 6.2, although there is not 100% repeatability due to different lighting and viewpoints, most of the interest points are robustly detected.

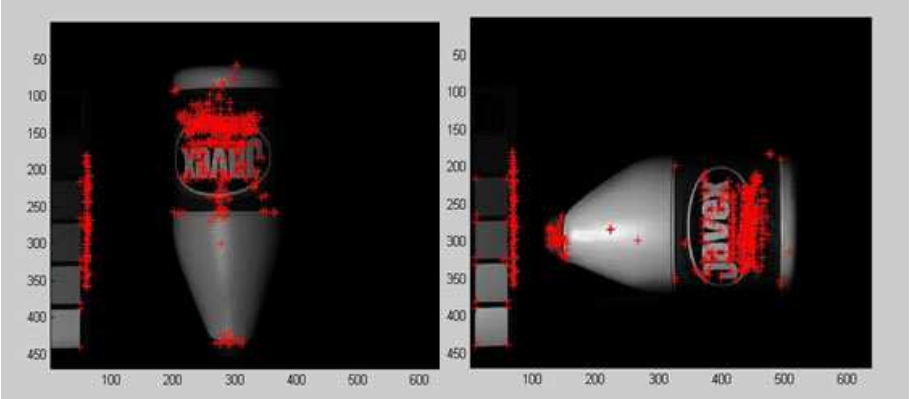

Figure 6.2: Interest points across illumination and rotation

As indicated in Fig. 6.2, using the whole image for calculations creates too much data for computation. Interest points are local features and occur where the signal changes two-dimensionally; where there is a corner or an edge or where the texture changes significantly. Interest points are rich in information content and are robust to many variables, including viewpoint, lighting, rotation and scaling [45].

6.2.2 Finding corresponding points from stitching images

The second task in the process is to find the partial correspondence between the two reference images. Here a new approach is used, based on local invariance, which is extended from content-based image retrieval research. This invariance characterises

neighbourhoods of interest points. Probably the most related work is that of Schmid and Mohr [63], who have investigated the matching of points between images from local grey level invariants and have also applied this to the retrieval of images from a database. However, their system only deals with invariance under rotation, combined with scale space, to overcome changes in scale between the images and the permitted change in viewpoint and light is rather limited. This work has extended these ideas toward invariance under more general transformations. More precisely, the work considers invariance under affine geometric transformations and under linear changes in intensities in each of the three colour bands, i.e. intensities change by a scale factor and offset that may be different for each colour band [131, 132]. To find corresponding interest points the invariance has to be found. The image in the neighbourhood of a point can be described by the set of its derivatives and one model is the "local jet model" which is derived using the convolution of the Gaussian derivatives. If I is an image and σ a given scale. The "local jet" of order N at a point $x = (x_1, x_2)$ is defined by

$$J^N[I](x,\sigma) = \left\{ L_{i_1 \ldots i_n}(x,\sigma) \,|\, (x,\sigma) \in IXIR^+; n = 0, \ldots, N \right\} \tag{6.4}$$

Where $L_{i_1 \ldots i_n}(x,\sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1 \ldots i_n}(x,\sigma)$ and $i_k \in \{x_1, x_2\}$. The σ of the Gaussian function is the quantity of the smoothing. The local jet is used to compute invariant feature vectors in individual interest points.

Previous work has applied invariant features such as local jets and neighbours, where the points and their constellation angles have to match [62, 63]. As indicated [62], not all neighbours have to be matched correctly, which is difficult when automatically finding corresponding points. Particularly the angular match is difficult for viewpoint variation. This new approach is to relax the constraints by using local invariant binary patterns (LBPs) which are used to describe the spatial structure of the local interest points. The binary values of the thresholded neighbourhood (Fig. 6.3a)

are mapped into an 8-bit word in clockwise or counter-clockwise order (Fig. 6.3b). In Fig. 6.3a, the corner response R values Rx from eq. 6.2 in interest point X and its eight neighbour interest points (Ra, Rb, Rc, Rd, Re, Rf, Rg, Rh) were compared. If $Ra >= Rx$, $A = 1$; otherwise $A = 0$. If $Rb >= Rx$, $B = 1$; otherwise $B = 0$. In this way, a binary number $ABCDEFGH$ can be obtained. An arbitrary number of binary shifts is then made (Fig. 6.3c), until the binary value is minimized and it is then used for indexing the spatial structure of the local interest points. The index of the matching pattern is used as the feature value, which is evaluated from Ojala's work [133], which describes the rotation-invariant LBP of the particular neighbourhood. Based on the local jet feature vectors and interest points neighbouring index using invariant LBP, corresponding points can be found. Fig. 6.4 illustrates the corresponding points from two stitching images where the two images were captured on different days where lighting, viewpoints and scales were all varied.

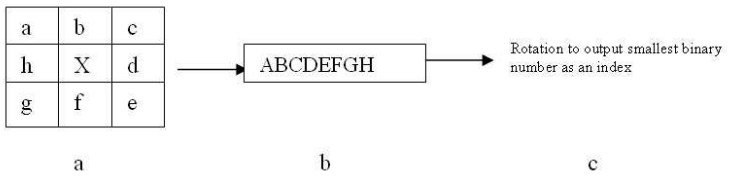

Figure 6.3: Rotation-invariant LBP

6.2.3 Spatial and Spectral Transform Matrices

This approach includes not only a spatial transform but also a spectral transform and the general image transform can be formulated as $IR = \{G, M, I1, I2\}$ where G is a planar projective transformation from the pixels of image $I1$ to that of image $I2$, in terms of homogeneous coordinates, and M is the spectral transform from the pixels of image $I1$ to that of image $I2$ using diagonal versus affine transformations for colour

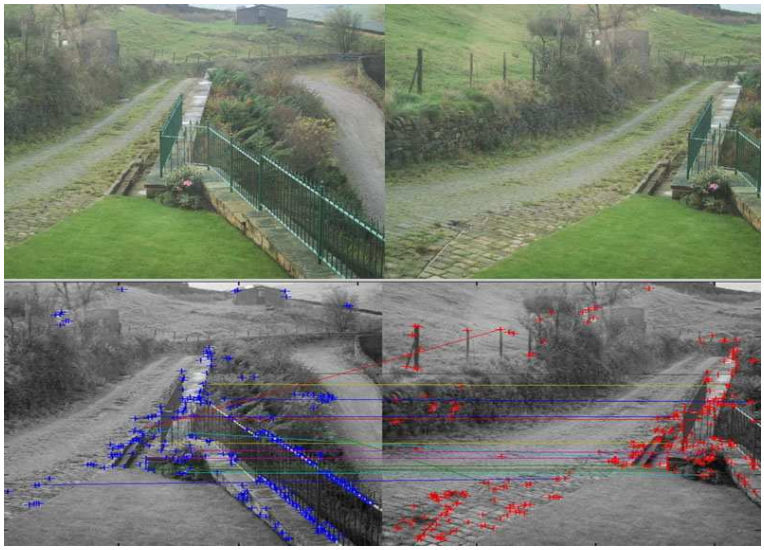

Figure 6.4: Two stitching images and their corresponding points

correction.

Spatial transform using homographic model

A non-singular linear transformation of the projective plane onto itself is called homography. The most general homography is represented by a 3x3 matrix G:

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} \\ H_{2,1} & H_{2,2} & H_{2,3} \\ H_{3,1} & H_{3,2} & H_{3,3} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6.5}$$

If we express points in homogeneous coordinates, then 2D points in the image plane are denoted as (x, y, z), with their corresponding Cartesian points being (x/z, y/z). If $H_{3,3} = 1$, then the matrix has 8 degrees of freedom, and $H_{3,3}$ is the scale factor. At least four points, which give 8 parameters, are needed to define a unique homography.

$$x' \left(H_{3,1}.x + H_{3,2}.y + H_{3,3}\right) = H_{1,1}.x + H_{1,2}.y + H_{1,3}$$

$$y'\left(H_{3,1}.x + H_{3,2}.y + H_{3,3}\right) = H_{2,1}.x + H_{2,2}.y + H_{2,3}$$

There are more than four points from the corresponding points found in the corresponding point tests from stitching images. All of the corresponding points are used to calculate homographic matrices. The 'best match' is found by using least square estimation. Homography is used to transform an image from one plane to another, and can be used to match images taken from different viewpoints. The images from Fig. 6.4 have been transformed using the following matrix and the stitched image, without colour correction, is displayed in Fig. 6.5.

$$G = \begin{bmatrix} 0.9299 & -0.0120 & 244.0661 \\ 0.0126 & 0.9886 & -0.06320 \\ -0.0000 & -0.0001 & 1.0000 \end{bmatrix}$$

Figure 6.5: Stitched image without colour correction

Spectral transform using the diagonal model

To compensate for differences in colour, a diagonal model [56, 130] is used. The light reflected from a surface depends on the spectral properties of the surface reflectance

and of the illumination incident on the surfaces:

$$\underline{\rho_k} = \int_{\omega} E(\lambda)\, S(\lambda) F_k(\lambda)\, d\lambda \tag{6.6}$$

Where λ is wavelength, ρ is the k-vector of sensor response (RGB pixel values), $F(\lambda)$ is the k-vector of sensor spectral sensitivities (the red, the green and the blue sensing channels), $E(\lambda)$ is the spectral power distribution of the illumination (assumed constant across the scene) , $S(\lambda)$ is the spectral reflectance function for a surface. The integral range ω is over the visible spectrum $\omega \in [380, 780nm]$. Let us approximate the spectral reflectance:

$$S(\lambda) \cong \sum_{j=1}^{3} \sigma_j s_j(\lambda) \tag{6.7}$$

Where $s_j(\lambda)$ are a set of n fixed basis functions. Let ρ denote the column vector of sensor measurement $\rho = (R, G, B)^T$, and let $\sigma = (\sigma_1 \sigma_2 \sigma_3)^T$ of spectral reflectance function coefficients. We can write

$$\underline{\rho}^E = \Lambda^E \underline{\sigma} \tag{6.8}$$

Where Λ^E is the 3x3 matrix with entries:

$$\Lambda^E = \int_{\omega} E(\lambda)\, s_j(\lambda) F_k(\lambda)\, d\lambda \tag{6.9}$$

The colours of a single surface viewed under different illuminants $E(\lambda)$ and eq. 6.8 is described by:

$$\underline{\rho}^E \cong \Lambda^E \underline{\sigma}$$

$$\underline{\rho}^e \cong \Lambda^e \underline{\sigma}$$

The vector σ describes the distribution of spectral reflectance of the object, and is independent of illumination. Because camera response is a linear transform of surface weight vector σ, it follows that camera responses are related across illumination by a linear transform:

$$\underline{\rho}^E \cong M \underline{\rho}^e$$

$$M = \Lambda^E \left[\Lambda^e\right]^{-1} \tag{6.10}$$

Eq. 6.8 plays a central role in this study of colour correction, since it effectively places a lower bound on the complexity of the illuminant colour problem. However, nine dimensional problems are hard to visualize and solve, particularly in real-time implementations. To simplify matters M in eq. 6.10 is often taken to be a 3-parameter diagonal matrix; thus reducing colour correction to a 3-parameter problem [130]. However, in this colour correction algorithm, the grey-world algorithms are expanded to diagonal plus affine transform [56]. All the pixels from the corresponding areas which were surrounded by the corresponding points in each image are used to calculate the spectral transform between the two images. In the eq. 6.11, the transform parameters can be obtained by 1st order polynomial fitting.

$$M = \begin{bmatrix} \alpha & & \alpha_1 \\ & \beta & \beta_1 \\ & \gamma & \gamma_1 \end{bmatrix} \tag{6.11}$$

Based on the above discussion, the transform matrix for the two images in Fig. 6.4 is shown as follows. The colour corrected stitched image is shown in Fig. 6.6.

$$M = \begin{bmatrix} 0.5908 & & & 32.2709 \\ & 0.6718 & & 24.2327 \\ & & 0.5152 & 40.8783 \end{bmatrix}$$

6.2.4 Image Mosaic with Smoothing

The final stage of the mosaic is to use the transform matrices to join the images, and a smoothing algorithm to 'fill' the 'holes' created after transformation due to the variations in image capture conditions, particularly variations in scaling. After undergoing geometric transformation and spectral transformation, a blending algorithm by using interest points preserving smoothing filter, is then applied to eliminate visible

Figure 6.6: The stitched images with colour correction

seams. The RGB values at any holes use linear interpolation from RGB values of its surrounding pixels. Again, the three RGB channels are processed independently. After individual processing, three grey images for R, G, B channels can simply be integrated together for colour image rendering. A smoothed image is illustrated in Fig. 6.6.

6.3 Experimental Tests

In order to provide accurate data for the homographic model to calculate spatial transform matrices it is important that the matches are as accurate as possible. One of the most popular interest point based matching algorithms is scale-invariant feature transform (SIFT), first proposed by Lowe in [49] and updated in [51]. For a more detailed look at SIFT refer back to section 2.3.5. The proposed interest point based system is compared with SIFT in a range of experiments. SIFT interest points (or keypoints) are invariant to changes in scale and rotation and robust to changes in viewpoint and illumination. To compare the two methods multiple 2-image tests were designed. The images used are from a diverse range of scenes, including outdoor scenes, office scenes and synthetic scenes involving a range of soft edge and hard edge

and high texture and low texture objects. The images were subject to a range of transformations, including translation (rotation about the optical axis of the camera), rotation (roll, negligible translation) and scaling (zoom, negligible translation). Some of the transformations were from data capture and some altered later in post processing. Because the transformations were either synthetic or easily measured (i.e. translation) it was possible to measure the positional accuracy. For each image pair the proposed and SIFT methods were both tested. The experimental results concentrated on how many points were detected and of those points, how many were an accurate match. The speed of the system was also measured. The tests used 50 images under translation, 20 under rotation and 10 under scale changes. The emphasis was placed on translation as this is the most common variable in panoramic imaging. A selection of the images used in the experiments is shown in Fig. 6.7.

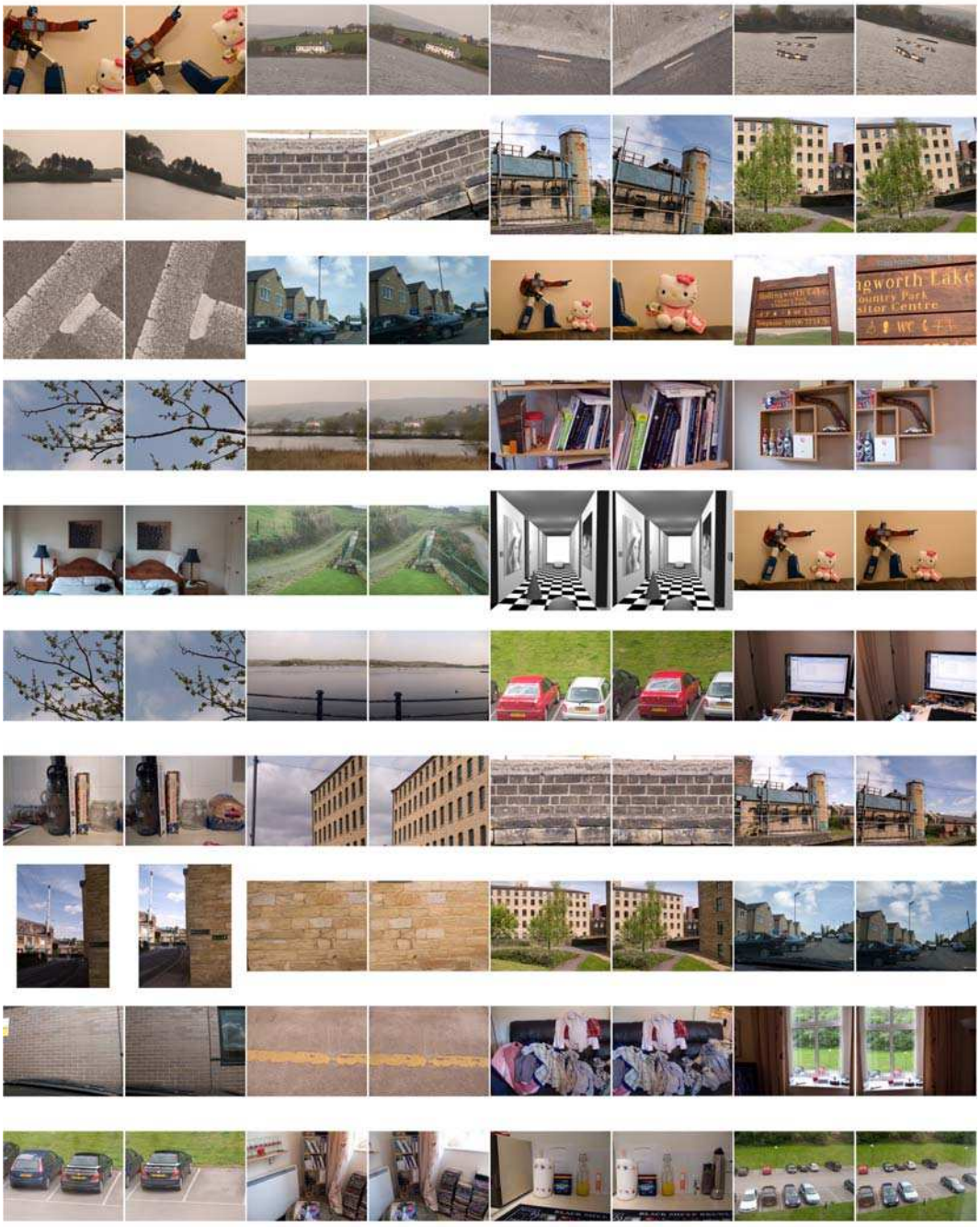

Figure 6.7: A selection of images used in the experimental tests for the feature matching algorithms

Averages for the accuracy and speed of the algorithms					
Image group	Method	# of matches	# of positive matches	% Accuracy	Time(s)
Translation	SIFT	1304	1143	85.20%	35.38
Translation	Proposed	101	60	75.32%	1.54
Rotation	SIFT	1574	1384	86.19%	37.80
Rotation	Proposed	100	77	79.29%	1.56
Scale	SIFT	606	367	50.88%	31.39
Scale	Proposed	50	6	5.71%	1.55
All images	SIFT	1285	1106	81.15%	35.39
All images	Proposed	95	57	67.61%	1.55

Table 6.1: Average values for the interest point correspondence results (SIFT vs. Proposed Method)

Table 6.1 shows the results from the experiments using the SIFT algorithm and the proposed algorithm. The results show that the SIFT algorithm tends to match many more features points than the proposed system, however the accuracy of those matches is only marginally better than the proposed system. The proposed algorithm does not match the accuracy of the SIFT algorithm, however it is significantly faster.

The results are more apparent in Fig. 6.8 showing higher average matching accuracy from the SIFT algorithm, most dramatic in scale changes. However, Fig. 6.9 shows the significant average speed increases from the proposed algorithm. In situations where only the translation or rotation are variables the proposed system could lend itself to near real time applications.

The minimum accuracy shown in Fig. 6.10 shows that the SIFT algorithm always has at least a few matching points to pass to a mosaicing application. The proposed system, under scale changes, would fail to give the mosaicing application accurate

Minimums for the accuracy and speed of the algorithms					
Image group	Method	# of matches	# of positive matches	% Accuracy	Time(s)
Translation	SIFT	173	83	47.98%	2.36
Translation	Proposed	3	3	37.75%	0.83
Rotation	SIFT	508	458	63.59%	9.77
Rotation	Proposed	23	15	64.15%	1.09
Scale	SIFT	25	3	12.00%	16.19
Scale	Proposed	1	0	0.00%	1.36
All images	SIFT	25	3	12.00%	2.36
All images	Proposed	1	0	0.00%	0.83

Table 6.2: Minimum values for the interest point correspondence results (SIFT vs. Proposed Method)

matching information. The minimum speed shown in Fig. 6.11 shows that under all circumstances the proposed algorithm is quicker. However, speed becomes irrelevant if no matching points are found.

The maximum accuracy shown in Fig. 6.12 suggests that under translation and rotation the proposed algorithm is capable of similar results to the SIFT algorithm. The maximum speed results (Fig. 6.13) also show that under no circumstances does the proposed algorithm take more than a couple of seconds, where as the SIFT algorithm can be processing for up to 56 seconds.

The standard deviation for the accuracy (Fig. 6.14) suggests that the SIFT algorithm is more stable than the proposed algorithm in general. However, this is not the case in the stability of the speed of the algorithm. The results show (Fig. 6.15) that the proposed algorithm is very consistent in the speed of the results, lending itself to a real time system, whereas the SIFT algorithm is very unstable, varying in speed

Maximums for the accuracy and speed of the algorithms					
Image group	Method	# of matches	# of positive matches	% Accuracy	Time(s)
Translation	SIFT	2747	2708	98.58%	56.68
Translation	Proposed	414	185	100.00%	2.12
Rotation	SIFT	3016	2763	93.84%	56.41
Rotation	Proposed	247	186	97.06%	1.79
Scale	SIFT	829	624	76.75%	39.03
Scale	Proposed	140	27	27.84%	1.93
All images	SIFT	3016	2763	98.58%	56.68
All images	Proposed	414	186	100.00%	2.12

Table 6.3: Maximum values for the interest point correspondence results (SIFT vs. Proposed Method)

from as fast as 2.36 seconds to as slow as 56.68 seconds.

In conclusion, the results show that the SIFT algorithm is more accurate than the proposed system in the majority of tests, however the proposed system is much quicker with only a minor accuracy penalty (except under scale change), lending itself more toward real time implementation.

Standard deviation for the accuracy and speed of the algorithms					
Image group	Method	# of matches	# of positive matches	% Accuracy	Time(s)
Translation	SIFT	694	662	11.34%	14.14
Translation	Proposed	121	55	19.43%	0.25
Rotation	SIFT	941	878	8.98%	14.05
Rotation	Proposed	73	53	10.03%	0.19
Scale	SIFT	333	227	23.97%	8.87
Scale	Proposed	64	12	12.37%	0.24
All images	SIFT	770	737	17.07%	13.42
All images	Proposed	105	55	28.90%	0.23

Table 6.4: Standard deviation values for the interest point correspondence results (SIFT vs. Proposed Method)

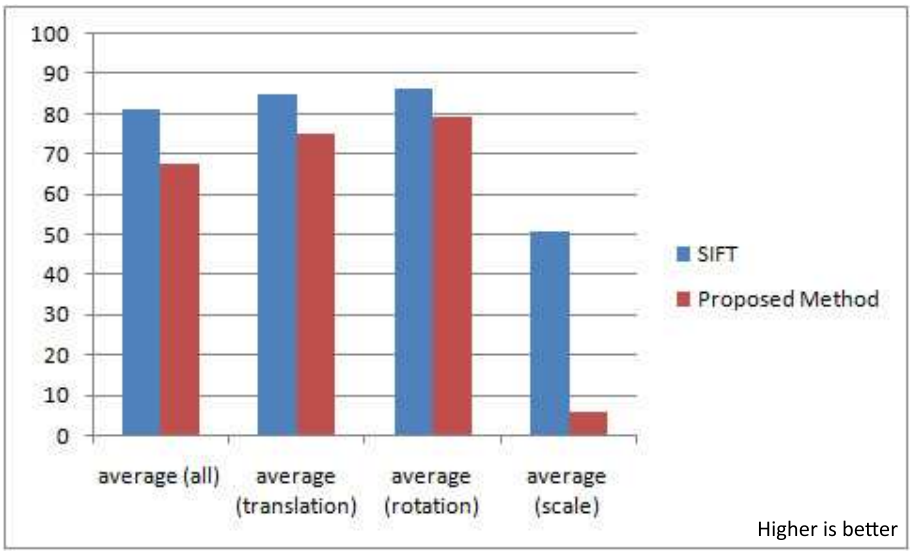

Figure 6.8: A comparison of the average accuracy (%) of the feature matching algorithms

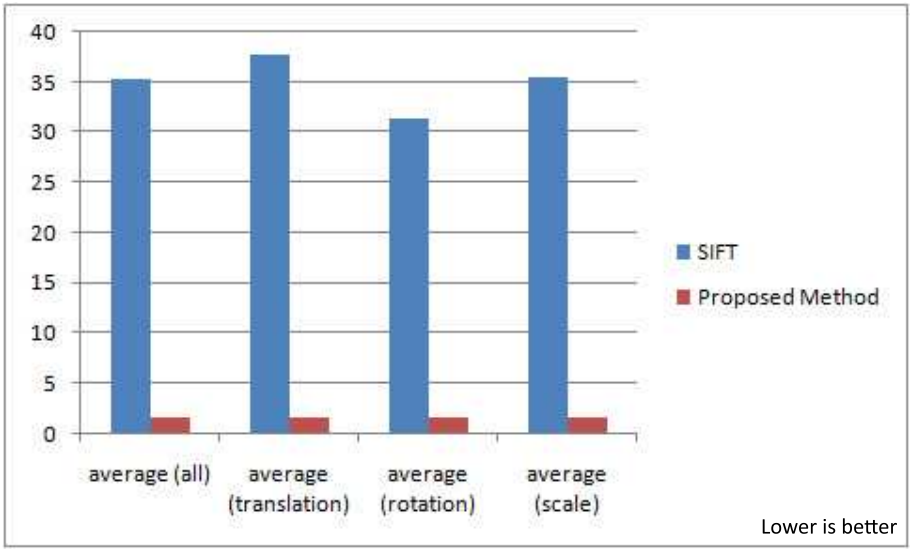

Figure 6.9: A comparison of the average speed (seconds) of the feature matching algorithms

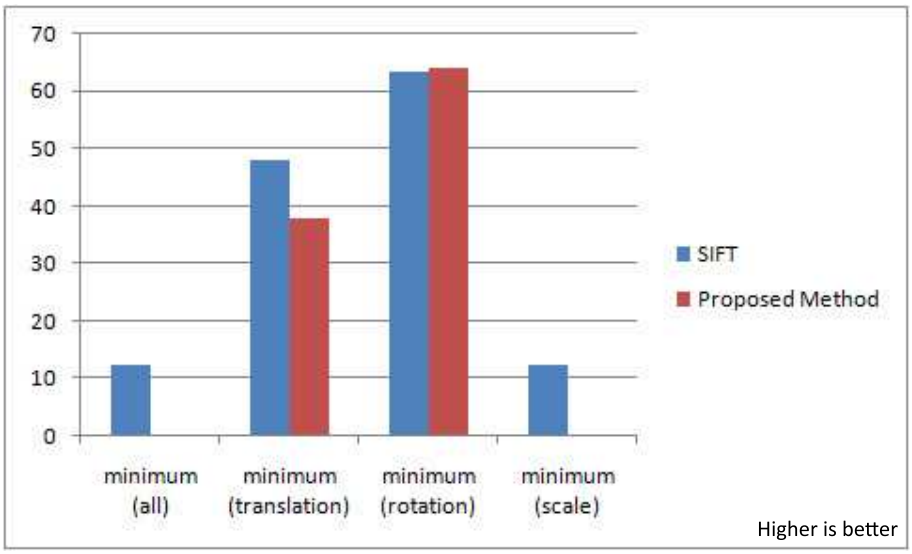

Figure 6.10: A comparison of the minimum accuracy (%) of the feature matching algorithms

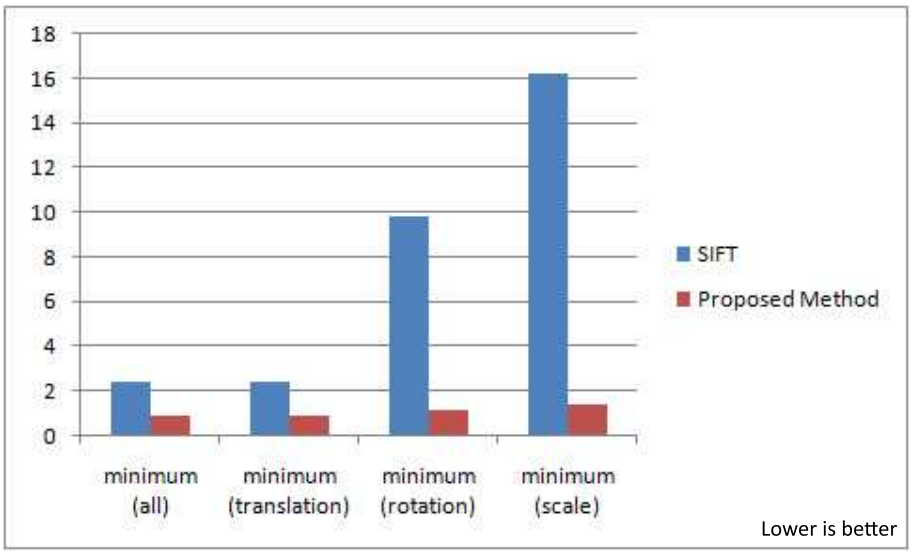

Figure 6.11: A comparison of the minimum speed (seconds) of the feature matching algorithms

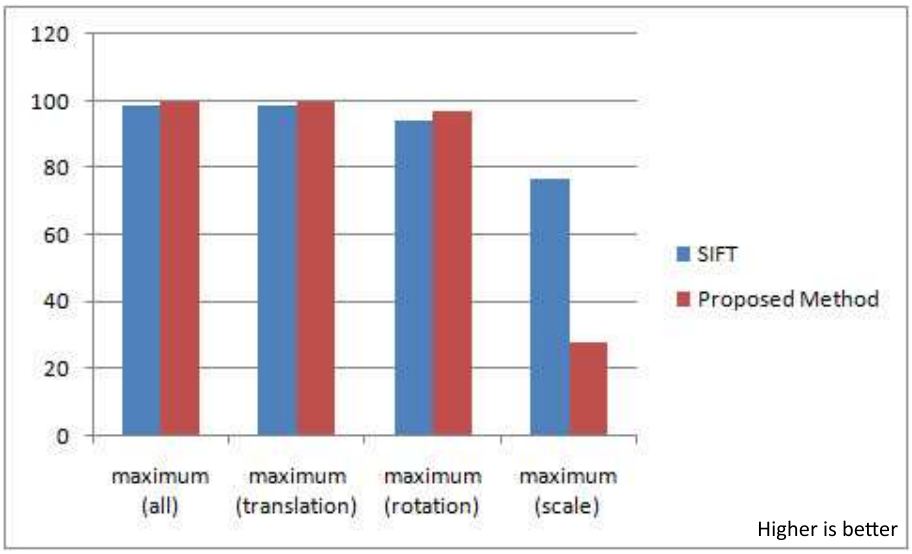

Figure 6.12: A comparison of the maximum accuracy (%) of the feature matching algorithms

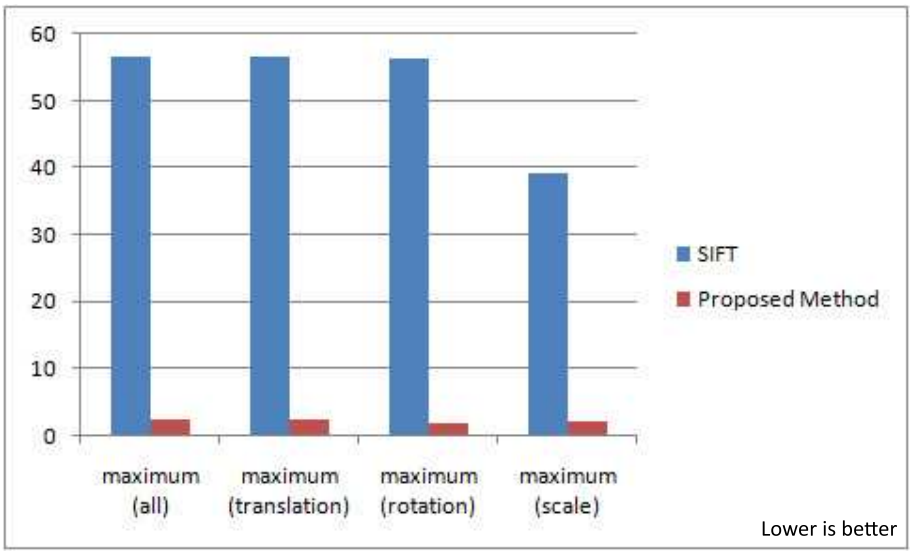

Figure 6.13: A comparison of the maximum speed (seconds) of the feature matching algorithms

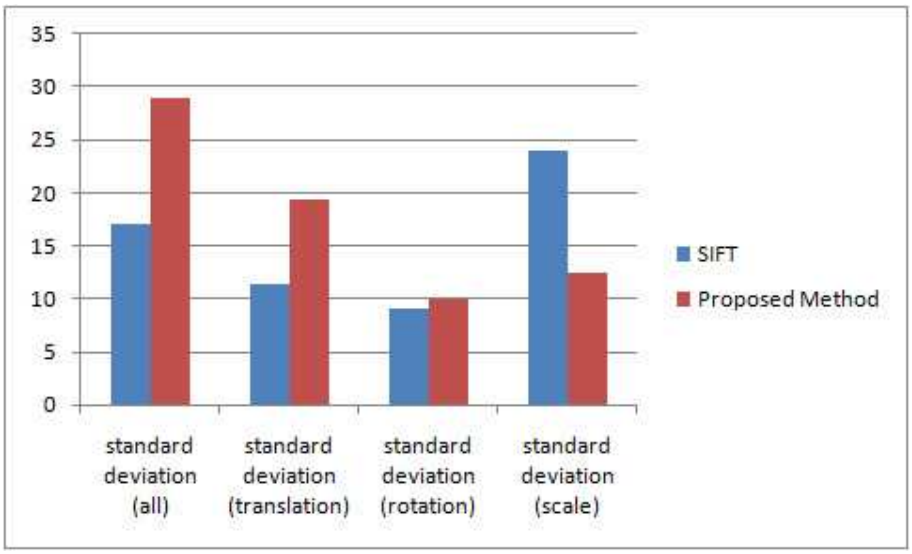

Figure 6.14: A comparison of the standard deviation of accuracy (%) of the feature matching algorithms

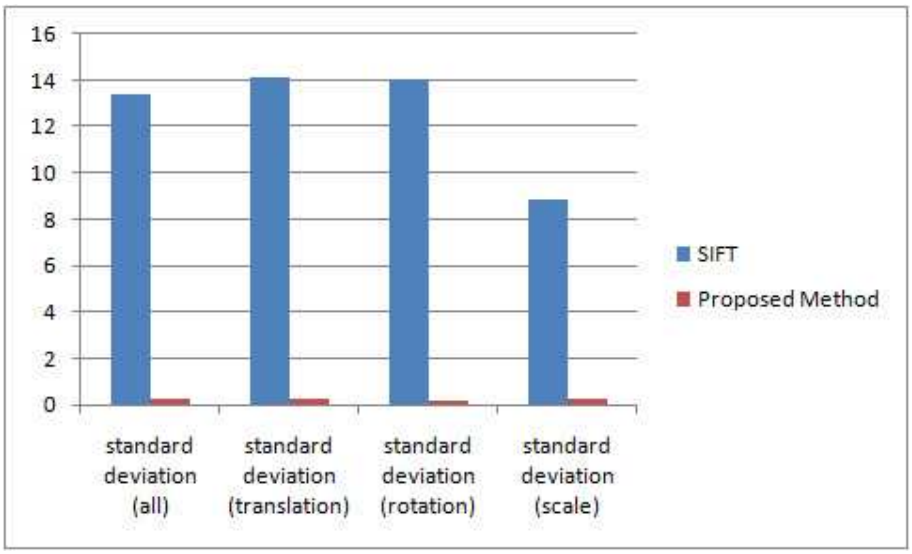

Figure 6.15: A comparison of the standard deviation of speed (seconds) of the feature matching algorithms

6.4 Summary

After reviewing the state-of-the-art in automatic image mosaic, a new approach has been described and validated. The new approach includes four steps of identifying interest points, finding corresponding points based on invariant features, geometrical and spectral transform and mapping. The preliminary results have demonstrated that the approach by using evaluated interest points and integration of geometrical & spectral transform is robust to various image-capturing conditions, although the system is not robust to large changes in scale. Compared with the SIFT algorithm, although the correspondence search is less accurate the proposed system is much quicker. Further work will expand the approach for other colour spaces and real-time searching for corresponding points, finding transform information and non-linear transforms among the corresponding points and apply more robust correspondence information [134, 135].

Chapter 7

3D SURFACE RECONSTRUCTION

7.1 Introduction

3D surface reconstruction is a popular problem. 3D surface reconstruction is required if the work is furthered to build 3D environments from panoramic images. Two processes to increase the accuracy and speed of 3D reconstruction are presented here. The first proposed process uses structured light to increase the accuracy of the correspondence search in low texture areas, e.g. white walls.

The second proposed approach involves a hybrid area based and feature based approach, using features to limit the area search to increase the speed of 3D reconstruction. This new proposed approach increases the speed of an area based method by between 30 and 50% without affecting the accuracy of the data.

7.2 3D Surface Reconstruction Using Structured Light

Correspondence methods are less accurate in areas of low texture. For example in an outdoor environment where texture is in abundance correspondence is more accurate, but an indoor environment usually has walls, and indoor walls usually have low texture, e.g. white paint. To overcome this lack of texture, it is proposed that a light pattern is projected onto the low texture areas to aid the correspondence search. Once a texture has been applied the correspondence algorithms achieve higher accuracy results. For this system a Gaussian noise pattern is produced. The image is filtered to ensure that no two dark pixels are next to each other, so that no "blocks" of black are produced. Large areas of black result in inaccurate removal of the noise

later. The structured light pattern has to be dense enough to create a usable texture for the correspondence algorithm, but with small enough 'dots' to be able to remove them for visualisation. The result is then projected into the environment. Fig. 7.1 shows an example of the structured light pattern used for projection. Because the pattern of structured light is known, a filter such as a median filter can be used to remove the pattern.

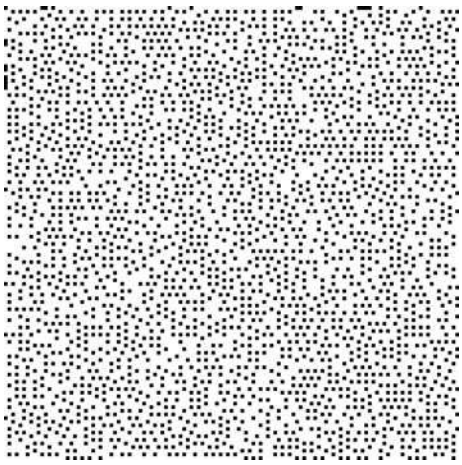

Figure 7.1: An example of the structured light pattern to be projected onto low texture surface areas

Once the 3D model has been created a texture needs to be applied if realism is to be achieved. The texture needed for the 3D model needs to be extracted from the captured image data, which has a structured light pattern projected onto it. For a realistic texture pattern the noise has to be removed. Median filtering techniques are used to remove the pattern from the image. Fig. 7.2 shows an example of the structured light pattern being removed from a wall image. Fig 7.3 shows the results of 3D depth calculations when no structured light pattern is used and Fig 7.4 shows the results when a pattern is used. The noise pattern projection for increased correspondence accuracy has also been used for 3D face reconstruction.

An example of the increase in accuracy gained by using the structured light pattern

Figure 7.2: Showing the structured light pattern in an image (left) and after it has been removed (right)

is shown in Fig. 7.5. The area within the red line clearly shows less disparity data noise.

7.3 Interest Point Based 3D Surface Reconstruction

To reduce search ambiguity due to occlusions, large disparities, photometric distortions and dense features and search time from large search area, many search windowing approaches have been proposed such as adaptive windows and multiple windows [112, 113, 114]. In reference to the divide-and-conquer method for Fast Fourier Transform (FFT) algorithms, this work expands the approach for 3D surface reconstruction [95]. Interest points are locations in the image where the signal changes two dimensionally. For example at a corner, a t-junction or where a significant texture change occurs. Many different interest point detectors exist. Interest points have a high repeatability rate, which means they are robust to changes in viewpoint, lighting, scale and rotation. The Harris detector is a very good interest point detector [49] so this chapter uses that method. Fig. 7.6 shows an example of an image with the interest points visible, and lines showing correspondence.

The area-based methods of finding stereo information are a slow search method as they find depth information pixel by pixel. They do however produce dense disparity

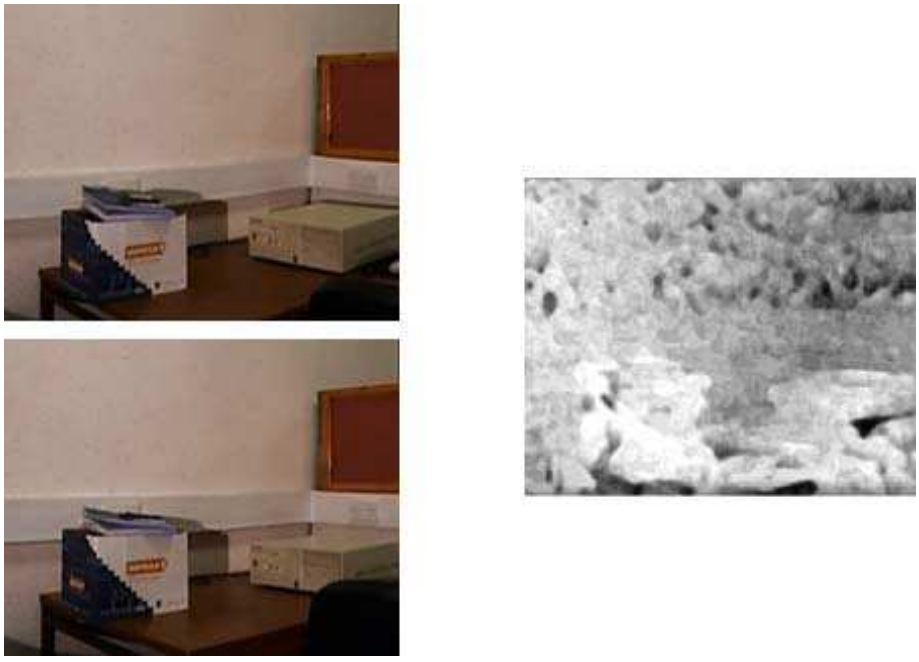

Figure 7.3: Showing the results of 3D depth calculations when no structured light pattern used

maps. As illustrated in Fig. 7.7, as the search area becomes larger, the computation time for exhaustive searching increases exponentially. Feature based methods are quicker but produce sparse disparity maps. This work proposes a mixture of these techniques so that a quicker method is available but which still creates a dense disparity map. The work also proposes a constraint for the search area from interest points. Using interest points to determine smaller corresponding areas of information for the search algorithm to search within, the area search becomes smaller. For example interest points find the corners of a box. The search is constrained within this box making a much smaller search region, which is searched much faster. Sun [136] proposed a similar method in which the image is subdivided into rectangles based on a coarse disparity map calculation. This coarse disparity map is used to determine areas of similarity which are subdivided and a dense disparity method applied. The

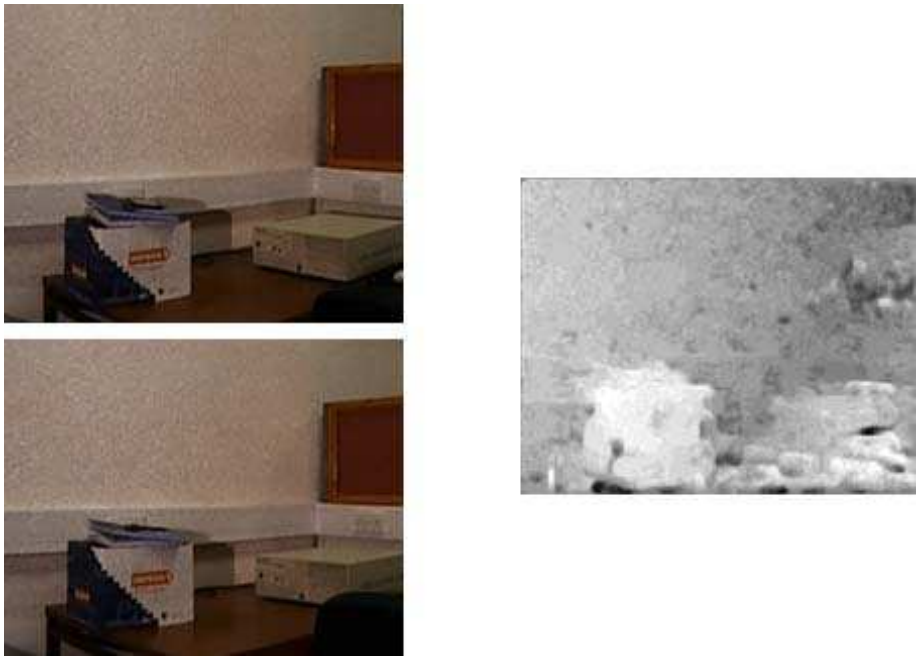

Figure 7.4: Showing the results of 3D depth calculations when structured light pattern used

difference in the proposed system is that the rectangles are chosen by interest point detection of regions rather than from a coarse disparity map. Interest points can be detected by using eq. 7.1:

$$\begin{aligned} \hat{C} &= \begin{vmatrix} \hat{I_x^2} & (\hat{I_x I_y}) \\ (\hat{I_x I_y}) & \hat{I_y^2} \end{vmatrix} \\ R(x,y) &= \det(\hat{C}) - k.trace^2(\hat{C}) \\ R(x,y) &> Threshold \Rightarrow Corner \end{aligned} \tag{7.1}$$

Where IX and IY are image intensity from horizontal and vertical direction Gaussian low pass filters. It is clear that different thresholds will detect different numbers of interest points. Fig. 7.7 shows a theoretical comparison curve of the number of calculations to size of search area using different size windows. It shows that the

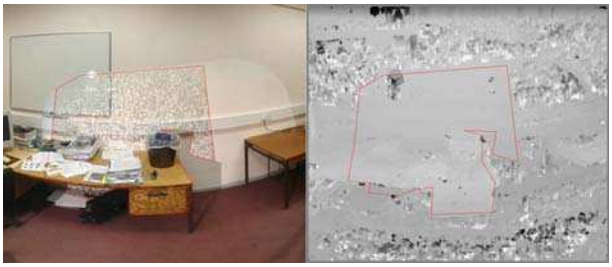

Figure 7.5: Structured light pattern projection results showing fewer errors in the depth map where the structured light is projected (inside the red line)

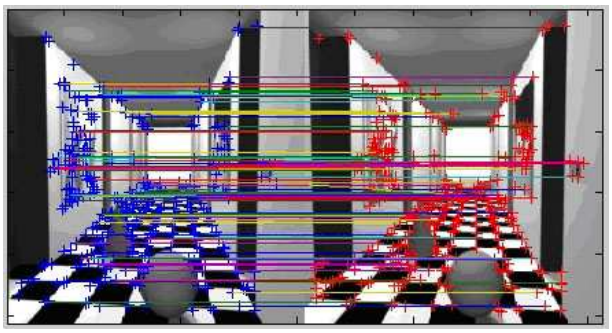

Figure 7.6: Interest points and their correspondence

smaller the search area is, the quicker it can be computed. Once the interest point data is available the image must be split into smaller areas, determined by the interest points. For example in Fig. 7.8 the interest points $I1$, $I2$, $I3$ and $I4$ have determined smaller areas. The corresponding points $I1'$, $I2'$, $I3'$ and $I4'$ in the second image would create a corresponding area for an area based correspondence search. In the implementation, the interest point based corresponding areas are overlapped.

Fig. 7.9 shows the process proposed. It shows the initial stages, using interest points to determine smaller areas of correspondence. These smaller areas are searched using area based methods for a dense disparity map. The disparity data is then combined to create a dense depth map of the whole scene.

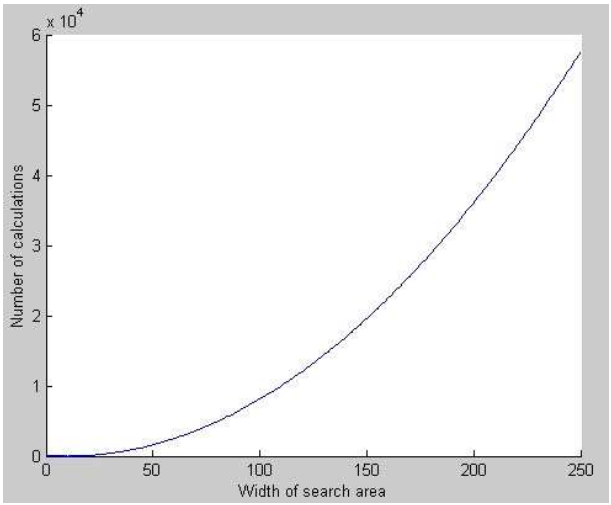

Figure 7.7: Computing expense and search area

7.4 Experimental Tests

Experimental tests were run on the images shown in Fig. 7.10. The images are corrected so that a search need only run on single rows. The image size is 256x256 pixels.

The images shown in Fig. 7.10 show the input stereo images and an output disparity map which was computed using an area based method. The computing time for this reconstruction was 391 seconds.

The method of finding interest points in the images and using that to determine smaller windows is much quicker. The time taken using the interest point method was approximately 196 seconds. The value is an estimate because the interest points do not cover the whole image. 51% of the data was processed in 100.9 seconds. This data shows a 49.9% time saving over the original method.

The calculations were also run on some brick data of rough surface reconstruction, shown in Fig. 7.11. The calculation time was reduced here from 632 seconds to 450 seconds. A time saving of 30% was shown in this experiment. The efficiency of time

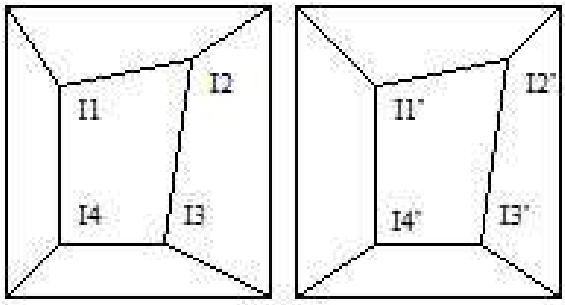

Figure 7.8: Corresponding areas for constrained area based searching (Left image and right image)

saving will be significantly increased for constructing large surface areas.

7.5 Summary

3D structure recovery of scenes is one of the most interesting goals of computer vision. The main approaches are structure-from-stereo, structure-from-motion and their integration. One of the most important processes for all the approaches is to find correspondences. A new approach based on interest points based 3D surface reconstruction for stereo vision has been proposed and tested. The computation can be reduced significantly for large 3D surface reconstruction. The approach has been tested for structure from stereo images with good results. It can be extended to structure from motion or their integration of stereo-motion. It can also be expanded to 3D panoramic construction. Further work would involve solving the problem of interest point coverage and overlaps of corresponding areas.

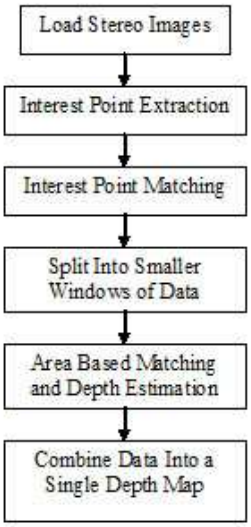

Figure 7.9: Process using interest points to increase the speed of processing the 2D data

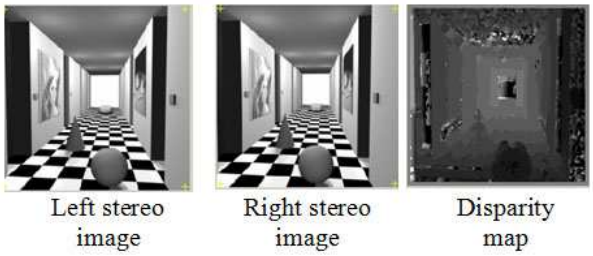

Figure 7.10: The stereo images and their disparity map, using area based method of correspondence searching

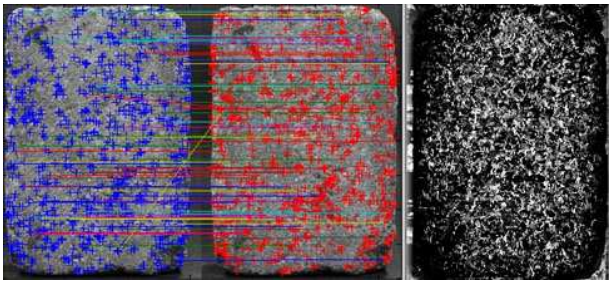

Figure 7.11: Brick data showing Interest points and correspondence matches, and disparity map

Chapter 8

CONCLUSIONS AND FURTHER WORK

Current work has concentrated on the capture and processing of multiple views for 3D virtual environmental construction. Based on the investigation and framework, which has been developed over the past few years, further work will be to extend the current work into building 3D environments from data captured with the 3D panoramic image capture system e.g. one-shot camera images. Applications in the real world will also be included.

8.1 Panoramic Imaging for Virtual Environment Creation

Firstly the panoramic image capture system will be used to capture panoramas from non-collinear positions in an environment. The captured data will then be processed using the proposed software system such as interest point detection and colour correction. Processing will include interest point detection in one-shot camera images, distortion correction, interest point correspondence matching, 3D surface reconstruction and visualisation in a virtual environment. A 3D video environment may be conceptualised and developed. Currently the system components are a manual process. The aim of the work beyond PhD level will not only be to create a 3D panoramic imaging system but also to automate the entire system. Real-time processing will also be considered. This would enable users to walk around an environment with a camera and a 3D virtual environment would be generated in real-time. Possible uses include wireless transmission and receiving of 3D environments, possibly to hand held computers. The background needed for future work now exists from the past few years

work, which will enable the work to move forward and create a real-time automated 3D panoramic system for the creation of virtual environments.

8.2 Web-Based Interactive System Using 3D Panoramic Imaging

In the future it will be possible to test real world applications including a web-based interactive system for business management, collaborative design and demonstrations (e-learning). An online collaborative system has been developed by a final year student at the University of Huddersfield. It would be possible that the proposed system could be used to build the 3D virtual building models used in the system. For example a user walking around a building site with a handheld camera capturing data. The model is built in the system offsite from the captured data and the collaborative design and maintenance decisions then shared across large networks. A proposed interactive system will be investigated as proposed in Fig. 8.1. More work on the interface and the image based 3D construction, storage and visualisation will be investigated.

Figure 8.1: Embedded 3D virtual environment for Web-based interactive applications

8.3 Video panoramic imaging

Research into a multiple camera real-time video panoramic system has begun. Using 6 small board cameras with an IEEE1394 interface aligned in a cylindrical layout it will be possible to capture video panoramas. It is the intention that the system might also be used to create stereo data. Eventually with two systems mounted above each other it will be possible to achieve real time stereo panoramic data at a higher resolution that the 'one-shot' system proposed. The data from the cameras has successfully been transformed into texture space. Mixing this with GPU's it is hoped a real-time system can be achieved.

8.4 Using interest points to integrate the real and virtual world in computer games

It is intended that through a new research grant the interest point work for stereo calculations will be extended into the augmented reality domain. The grant is to extend the work so that computer games can include virtual characters in real world environments. As well as computer games applications the work will also look at training and visualisation.

8.5 Conclusions

Achievements over the past few years have been accomplished in 3 main areas, colour correction, image stitching and 3D surface reconstruction. The work achieved will enable a 3D panoramic system to be developed in the future.

The major contributions can be summarised as follows:

- A comprehensive literature survey has been undertaken. The state of the art computer vision techniques and graphics techniques have been reviewed. Specifically methods for colour correction, interest points, panoramic imaging and 3D

surface reconstruction have been covered [137].

- A new interest point based image stitching method has been proposed and investigated. The robustness of interest points has been tested and evaluated. Interest points have been proved to be very suitable for 3D panoramic imaging as all these changes can occur [109]. Further work to implement them into an open source mosaicing application like Hugin or PTGui would be very interesting.

- A new interest point based method for colour correction has been proposed and investigated. The results of linear and linear plus affine colour transforms have proved more accurate than traditional diagonal transforms in accurately matching colours in panoramic images [137]. Using interest points to determine overlap regions and then mapping the spectral transform matrix has resulted in accurate colour correction, meaning a much more aesthetically pleasing result.

- A new interest point based 3D surface reconstruction method has been proposed and investigated. Traditional area based methods have been modified to include an initial interest point and correspondence search. Limits have then been placed on the area search resulting in increased speed performance. Increases of 30-50% have been recorded, without loss of detail in the data. Using structured light to increase the accuracy of the correspondence search has also been very successful.

- A new hardware rig has been setup for 3D panoramic testing. The hardware rig is capable of low resolution one-shot capture, limited in vertical field of view. The rig is also capable of high resolution multi-shot capture with no limits on field of view.

- A new software framework has been developed for image based 3D virtual environment construction. The GUI includes abilities for importing images, colour

correction, stitching, 3D surface reconstruction and visualisation.

- A comparative study using multiple views from normal digital cameras and one-shot camera images for 3D environment construction has been undertaken. The merits and disadvantages of these approaches have been identified, which will be the basis for further study.

BIBLIOGRAPHY

[1] Canon Inc. http://www.canon.com.

[2] Pentax. http://www.pentax.com.

[3] Panorama Factory. http://www.panoramafactory.com.

[4] Ipix. http://www.ipix.com.

[5] BeHere. http://www.behere.com.

[6] 0-360. http://www.0-360.com.

[7] Spheron. http://www.spheron.com.

[8] D Vision Works. http://www.d-vw.com.

[9] Canon. http://www.canon.co.uk/About_Us/About_Canon/Research__Development/3D_Softw

[10] Rotography Ltd. http://www.rotography.com.

[11] Helmut Dersch. http://www.fh-furtwangen.de/ dersch.

[12] S. E. Chen. "Quicktime VR - An Image Based Approach to Virtual Environment Navigation". *Proc. SIGGRAPH*, pages 29–38, 1995.

[13] Nikon Cameras. http://www.nikon.co.uk.

[14] Kodak Cameras. http://www.kodak.com.

[15] Kaidan products. http://www.kaidan.com/products/pano-prods.html.

[16] Agnos Equipment. http://www.agnos.com/.

[17] Pixaround. http://www.pixaround.com.

[18] MGI Webtools. http://www.mgisoft.com/products/webtools/panorama/index.asp.

[19] iMove Inc. http://www.imoveinc.com.

[20] D. W. Rees. "Panoramic Television Viewing System". *U.S. Patent No. 3,505,465*, April 1970.

[21] T. Svoboda, T. Pajdla, and V. Hlavac. "Central Panoramic Cameras: Design and Geometry". *Third Computer Vision Winter Workshop*, Feb. 1998.

[22] M. Aggarwal and N. Ahuja. "High Dynamic Range Panoramic Imaging". *Eighth IEEE International Conference on Computer Vision (ICCV2001)*, 1:2–9, 2001.

[23] H. Huang and Y. Hung. "Panoramic Stereo Imaging System with Automatic Disparity Warping and Seaming". *Graphical models and Image Processing*, 60(3):196–208, May 1998.

[24] H. Shum and R. Szeliski. "Stereo Reconstruction from Multiperspective Panoramas". *7th IEEE Int. Conference on Computer Vision*, Sept. 1999.

[25] S. Peleg and M. Ben-Ezra. "Stereo Panorama with a Single Camera". *CVPR99*, 1:1395, 1999.

[26] H Ishiguro, M Yamamoto, and S Tsuji. Omni-directional stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1992.

[27] D. Tzovaras, N. Grammalidis, and M. G. Strintzis. "Disparity Field and Depth Map Coding for Multiview 3D Image Generation". *Signal Processing: Image Communication 11*, pages 205–230, 1998.

[28] G. Medioni, C. Tang, and M. Lee. "Tensor Voting: Theory and Applications". *12eme Congres Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, 2000.

[29] P. M. Q. Aguiar and J. M. F. Moura. "Three-dimensional Modelling from two-dimensional Video". *IEEE Transactions on Image Processing*, 10(10):1541–1551, Oct. 2001.

[30] Z. Zhang. "A Flexible New Technique for Camera Calibration". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), Nov. 2000.

[31] J. Heikkila. "Geometric Camera Calibration using Circular Control Points". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, Oct. 2000.

[32] L. McMillan. "Acquiring Immersive Virtual Environments with an Uncalibrated Camera". Technical Report 95-006, University of North Carolina, 1995.

[33] Q. T. Luong and O. D. Faugeras. "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices". *International Journal of Computer Vision*, 22(3):261–289, 1997.

[34] K. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter. http://www.vision.caltech.edu/bouguetj/calib_doc/.

[35] F. Odone and A. Fusiello. "Applications of 2D Image Registration". Research Memorandum RM/99/15, Heriot-Watt University.

[36] L. G. Brown. "A Survey of Image Registration Techniques". *ACM Computing Surveys*, 24(4):325–376, Dec. 1992.

[37] C. L. Zitnick and T. Kanade. "A Cooperative Algorithm for Stereo Matching and Occlusion Detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684, July 2000.

[38] K. Yamazawa, Y. Yagi, and M. Yachida. "Obstacle Avoidance with Omnidirectional Image Sensor HyperOmni Vision". *Proc. of IEEE International Conference on Robotics and Automation*, pages 1062–1067, 1995.

[39] J. Canny. "A Computational Approach to Edge Detection". *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679–714, 1986.

[40] P. Bao and D. Xu. "Complex Wavelet-Based Image Mosaics using Edge-Preserving Visual Perception Modelling". *Computers and Graphics*, 23:309–321, 1999.

[41] H. S. Kim and C. H. Kim H. C. Kim, W. K. Lee. "Stitching reliability for estimating camera focal length in panoramic image mosaicing". *15th International Conference on Pattern recognition*, pages 596–599, 2000.

[42] E. Bourque and G. Dudek. "Automated Creation of Image Based Virtual Reality". *Proceedings of SPIE, "Sensor Fusion and Decentralised Control in Autonomous Robotic Systems"*, 3209:292–301, 1997.

[43] H. Moravec. "Rover Visual Obstacle Avoidance". *International Joint Conference on Artificial Intelligence*, 2:785–790, 1981.

[44] C. Harris and M. Stephens. "A combined corner and edge detector". *Fourth Alvey Vision Conference*, pages 147–151, 1988.

[45] C. Schmid, R. Mohr, and C. Bauckhage. "Evaluation of Interest Point Detectors". *International Journal of Computer Vision*, 37(2):151–172, 2000.

[46] I. Zoghlami, O. Faugeras, and R. Deriche. "Using geometric corners to build a 2D mosaic from a set of images". *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–425, 1997.

[47] K. Mikolajczyk and C. Schmid. "Indexing Based Scale Invariant Interest Points". *International Conference on Computer Vision*, pages 525–531, 2001.

[48] C. Schmid and R. Mohr. "Image Retrieval using Local Characterisation". *Proc. of the 3rd International Conference on Communicating by Image and Multimedia*, pages 45–50, May 1996.

[49] D. G. Lowe. "Object Recognition from Local Scale-Invariant Features". *Proceedings of the International Conference on Computer Vision*, pages 1150–1157, Sept. 1999.

[50] S. M. Smith and J. M. Brady. "SUSAN - A New Approach to Low Level Image Processing". *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[51] D. G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". *Int. Journal of Computer Vision*, 60:91–110, 2004.

[52] A. P. Witkin. "Scale-space filtering". *International Joint Conference on Artificial Intelligence*, pages 1019–1022, 1983.

[53] H. E. Ives. "The Relation Between the Color of the Illuminant and the Color of the Illuminated Object". *Trans. Illuminat. Eng. Soc.*, 7:62–72, 1912.

[54] E. H. Land and J. J. McCann. "Lightness and retinex theory". *Journal of the Optical Society of America A*, 61:1–11, 1971.

[55] G. Finlayson, M. Drew, and B. Funt. "Color Constancy: Generalized Diagonal Transforms Suffice". *Journal of the Optical Society of America A*, 11(11):3011–3020, 1994.

[56] B. V. Funt and B. C. Lewis. "Diagonal versus Affine Transformations for Colour Correction". *Journal of the Optical Society of America*.

[57] Apple Computer Inc. "An Overview of Apple's QuickTime VR Technology". Technical report, http://quicktime.apple.com/qtvr/qtvrtech5_25.html, 1995.

[58] R. Szeliski. "Image mosaicing for tele-reality applications". *WACV94*, pages 44–53, 1994.

[59] R. Szeliski. "Video mosaics for virtual environments". *IEEE Computer Graphics and Applications*, pages 22–30, 1996.

[60] R. Szeliski and H. Shum. "Creating Full View Panoramic Image Mosaics and Environment Maps". 1998.

[61] L. Zhao and Y. H. Yang. "Mosaic image method: a local and global method". *The IEICE Transactions Information and Systems*.

[62] C. Schmid and R. Mohr. "Local gray-value invariants for image retrieval". *IEEE Transactions on Pattern Analysis and Machine Intelligent*, 19(5):530–535, 1997.

[63] C. Schmid, R. Mohr, and C. Bauckhage. "Comparing and evaluating interest points". *Proc. of the International Conference on Computer Vision*, 1998.

[64] O. Faugeras. "Three Dimensional Computer Vision: a Geometric Viewpoint". *MIT Press*, 1993.

[65] H. Y. Shum and R. Szeliski. "Construction and Refinement of Panoramic Mosaics with Global and Local Alignment". *In Proc. IEEE Int. Conf. On Computer Vision*, pages 953–958, 1998.

[66] A. Can, C. Stewart, and B. Roysam. "Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina". *In Proc. CVPR*, pages 286–292, 1999.

[67] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry". *Artificial Intelligence*, 78(1-2):87–119, 1995.

[68] K. Kanatani and N. Ohta. "Accuracy Bounds and Optimal Computation of Homograph for Image Mosaicing Applications". *The Seventh International Conference on Computer Vision (ICCV'99)*, Sept. 1999.

[69] A. Fusiello. "Notes on the Applications of Homographies in Computer Vision". Research Memorandum RM/99/13, Department of Computing and Electrical Engineering, Heriot-Watt University, Edinburgh, UK, 1999.

[70] S. Mann and R. W. Picard. "Virtual Bellows: Contructing High Quality Stills from Video". Technical Report 259, M.I.T Media Laboratory Perceptual Computing Section, 1994.

[71] M. Kourogi, T. Kurata, J. Hoshino, and Y. Muraoka. "Real-time Image Mosaicing From a Video Sequence". *Proc. ICIP99*, 4:133–137, 1999.

[72] Z. Zhu, A. R. Hanson, H. Schultz, F. Stolle, and E. M. Riseman. "Stereo Mosaics from a Moving Video Camera for Environmental Monitoring". *First International Workshop on Digital and Computational Video*, pages 45–54, 1999.

[73] P. J. Burt and E. H. Adelson. "A Multiresolution Spline with Application to Image Mosaics". *ACM Transactions on Graphics*, 2(4):217–236, Oct. 1983.

[74] 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*, 71(2):153–165, Aug. 1998.

[75] H. C. Longuet-Higgins. "A computer algorithm for reconstructing a scene from two projections". *Nature*, 293:133–135, September 1981.

[76] U. R. Dhond and J. K. Aggarwal. "Structure from Stereo - A Review". *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6), 1989.

[77] S. Yu, S. R. Sukumar, A. F. Koschan, D. L. Page, and M. A. Abidi. 3D Reconstruction of Road Surfaces Using an Integrated Multi-Sensory Approach. *Optics and Lasers in Engineering*, 45:808–818, 2007.

[78] V. Gerdes. http://www-student.cs.uni-bonn.de/ gerdes/MRTStereo/index.html.

[79] J. Park and S. Inoue. "Hierarchical Depth Mapping from Multiple Cameras". *Proc. ICIAP'97*, 1:685–692, 1999.

[80] H.M. Hamden, E.E. Hemayed, and A. A. Farag. "A Fast 3D Object Reconstruction Using Trinocular Vision and Structured Light". *Proc. of SPIE*, Conf. 3522, Nov. 1999.

[81] R. J. Valkenburg. "Accurate 3D Measurement using a Structured Light System". *Image and Vision Computing*, 1996.

[82] M. Pilu. "A Direct Method for Stereo Correspondence Based on Singular Value Decomposition". *IEEE CVPR'97*, June 1997.

[83] C. Tsai and A. K. Katsaggelos. "Dense Disparity Estimation with a Divide-and-Conquer Disparity Space Image Technique". *IEEE Transactions on Multimedia*, 1(1), 1999.

[84] D. Jelinek and C. J. Taylor. "Reconstruction of Linearly Parameterized Models from Single Images with a Camera of Unknown Focal Length". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(7):767–773, July 2001.

[85] E. Izquierdo and S. Kruse. "Image Analysis for 3D Modelling, Rendering and Virtual View Generation". *Computer Vision and Image Understanding*, 71(2):231–253, Aug. 1998.

[86] W. Lee, J. Gu, and N. Magnenat-Thalmann. "Generating Animatable 3D Virtual Humans from Photographs". *Eurographics 2000*, 19(3), 2000.

[87] S. M. Seitz and C. R. Dyer. "Physically-valid View Synthesis by Image Interpolation". *Proc. Workshop on Representations of Visual Scenes*, 1995.

[88] P. E. Debevec, C. J. Taylor, and J. Malik. "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry and Image Based Approach". *Proc. ACM SIGGRAPH 96*, pages 11–20, 1996.

[89] B. M. Oh, M. Chen, J. Dorsey, and F. Durand. "Image-Based Modeling and Photo Editing". *Proc. ACM SIGGRAPH 2001*, pages 433–442, 2001.

[90] D. G. Aliaga and I. Carlbom. "Plenoptic Stitching: A Scalable Method for Reconstructing 3D Interactive Walkthroughs". *Proc. Of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 443–450, Aug. 2001.

[91] J. Park and S. Inoue. "Hierarchical depth mapping from multiple cameras". *Proc. ICIAP'97*, 1:685–692, Sept. 1997.

[92] O. Faugeras, B. Hotz, H. Mathieu, T. Vieville, Z. Zhang, P. Fua, E. Theron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy. "Real time correlation based stereo: algorithm implementations and applications". Technical Report 2013, I.N.R.I.A, 1993.

[93] H. M. Hamdan, E. E. Hemayed, and A. A. Farag. "A Fast 3D Object Reconstruction Using Trinocular Vision and Structured Light". *Proc. of SPIE, Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision*, 3522, Nov. 1998.

[94] H. H. Baker and T. O. Binford. "Depth from Edge and Intensity Based Stereo". *Int. Joint Conf. on Artificial Intelligence*, pages 631–636, 1981.

[95] C. P. Jerian and R. Jain. "Structure from Motion - a Critical Analysis of Methods". *IEEE Trans. Systems, Man, and Cybernetics*, 21(3):572–588, 1991.

[96] C. Tomasi and T. Kanade. "Shape and motion from image streams under orthography: a factorisation approach". *International Journal of Computer Vision*, 9(2):137–154, 1992.

[97] M. G. Mostafa, S. M. Yamany, and A. A. Farag. "Integrating Stereo and Shape from Shading". *IEEE International Conference on Image Processing (ICIP'99)*, 3:130–134, Oct. 1999.

[98] K. K. Wong and R. Cipolla. "Structure and motion from silhouettes". *ICCV01*, 2001.

[99] E. Trucco and A. Verri. "Introductory Techniques for 3D Computer Vision". *Prentice Hall*, 1998.

[100] L. Zhang. "Hierarchical Block-Based Disparity Estimation Using Mean Absolute Difference and Dynamic Programming". *VLBV01: International Workshop on Very Low Bitrate Video Coding*, pages 114–118, 2001.

[101] B. J. Super and W. N. Klarquish. "Patch-based Stereo in a General Binocular Viewing Geometry". *IEEE Trans. Pattern Analysis and Machine Vision*, 19(3):247–253, 1997.

[102] H. Shum, A. Kalai, and S. M. Sietz. "Omnivergent Stereo". *Proc. Seventh International Conference on Computer Vision (ICCV'99)*, 1999.

[103] R. I. Hartley and P. Sturm. "Triangulation". *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[104] A. N. Rajagopalan and S. Chaudhuri. "Optimal Recovery of Depth from Defocused Images Using an MRF Model". *ICCV98*, pages 1047–1052, 1998.

[105] M. Borga and H. Knutsson. "Estimating Multiple Depths in Semi-transparent Stereo Images". *Proceedings of the 11th Scandinavian Conference on Image Analysis*, June 1999.

[106] Q. T. Luong and O. D. Faugeras. "The Fundamental Matrix: Theory, Algorithms and Stability Analysis". *Int. Journal of Computer Vision*, 17:43–75, 1996.

[107] R. Oliveira, J. Xavier, and J. P. Costeira. Multi-View Correspondence by Enforcement of Rigidity Constraints. *Image and Vision Computing*, 25(6):1008–1020, June 2007.

[108] C. J. Taylor and D. Jelinek. "Reconstruction of Linearly Parameterized Models from Single Images with a Camera of Unknown Focal Length". *IEEE PAMI*, 23(7), 2001.

[109] G. Y. Tian, D. Gledhill, and D. Taylor. "Comprehensive interest points based imaging mosaic". *Pattern Recognition Letters*, 24:1171–1179, 2003.

[110] D. Bergmann and R. Ritter. "3D Deformation Measurement in Small Areas based on Grating Method and Photogrammetry". *European Symposium on Lasers, Optics and Vision for Productivity in Manufacturing*, 1996.

[111] W. Hoff and N. Ahuja. "Surface from stereo: integrating feature matching, disparity estimation and contour detection". *IEEE Transaction on Pattern Recognition and Machine Intelligence*, 11(2), 1989.

[112] T. Kanade and M. Okutomi. "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment". *IEEE Trans. Pattern Analysis and Machine Vision*, 16(9):920–932, 1994.

[113] A. Fusiello, V. Roberto, and E. Trucco. "Efficient Stereo with Multiple Windowing". *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 858–863, 1997.

[114] O. Veksler. "Stereo correspondence with compact windows via minimum ratio cycle". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1654–1660, 2002.

[115] N. Grammalidis and M. G. Strintzis. "Disparity and Occlusion Estimation in Multiocular Systems and Their Coding for the Communication of Multiview Image Sequences". *IEEE Transactions on Circuits and Systems for Video Technology*, 8(3), 1998.

[116] C. Malerczyk, K. Klein, and T. Wiebesiek. "3D reconstruction of sport events for digital TV". *Proceedings of WSCG*, 11(1), Feb. 2003.

[117] C. M. Kuo, C. P. Chao, and C. H. Hsieh. "An efficient motion estimation algorithm for video coding using kalman filter". *Real-time imaging*, 8:253–264, 2002.

[118] J. Weng, T. S. Huang, and N. Ahuja. "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis and Error Estimation". *IEEE Trans. Pattern Analysis and Machine Vision*, 11(5):451–476, 1989.

[119] F. Domaika and R. Chung. "Binocular matching constraints from motion". *Pattern Recognition*, 35:2003–2012, 2002.

[120] Y. Pritch, M. Ben-Ezra, and S. Peleg. "Optics for Omnistereo Imaging". *Foundations of Image Understanding*, pages 447–467, July 2001.

[121] J. Gluckman, S. K. Nayar, and K. Thorek. "Real-Time Omnidirectional and Panoramic Stereo". *Image Understanding Workshop (IUW)*, 1998.

[122] T. Svoboda and T. Pajdla. "Panoramic Cameras for 3D Computation". *Proc. Of the Czech Pattern Recognition Workshop, Czech Society for Pattern Recognition*, pages 63–70, Feb. 2000.

[123] H. Koyasu, J. Miura, and Y. Shirai. "Real-time omnidirectional stereo for obstacle detection and tracking in dynamic environments". *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1:31–36, 2001.

[124] R. Szeliski and H. Shum. "Creating Full View Panoramic Image Mosaics and Environment Maps". *Proc. SIGGRAPH 97*, pages 251–258, 1997.

[125] K. Lee, Y. Fung, K. Wong, S. Or, and T. Lao. "Panoramic Video Representation using Mosaic Image". *Proc. of CISST'99*, pages 390–396, June 1999.

[126] X. Sun, J. Foote, D. Kimber, and B. S. Manjunath. "Panoramic video capturing and compressed domain virtual camera control". *Proc. 9th ACM International Conference on Multimedia*, pages 329–347, 2001.

[127] H. Shum and S. B. Kang. "A Review of Image-based Rendering Techniques". *IEEE/SPIE Visual Communications and Image Processing (VCIP)*, pages 2–13, June 2000.

[128] K. Kiyokawa, Y. Kurata, and H. Ohno. "An Optical See-through Display for Mutual Occlusion with a Real-time Stereovision System". *Computers and Graphics 25*, pages 765–779, 2001.

[129] A. D. Griffiths, A. J. Coates, J. L. Josset, G. Paar, B. Hofman, D. Pullan, P. Ruffer, M. R. Sims, and C. T. Pillinger. The Beagle 2 stereo camera system. *Planetary and Space Science*, 53:1466–1482, 2005.

[130] G. D. Finlayson and G. Y. Tian. "Colour normalisation for colour object recognition". *International J. of Pattern Recognition and Artificial Intelligence*, 13(8):1271–1285, 1999.

[131] G. Healey and D. Slater. "Global color constancy: recognition of objects by use of illumination invariant properties of color distributions". *Journal of the Optical Society of America A*, 11(11):3003–3010, 1995.

[132] T. Tuytelaars and L. Van Gool. "Content-based image retrieval based on local affinity invariant regions". *Proceedings Visual '99: Information and Information Systems*, pages 493–500, 1999.

[133] T. Ojala. *"Nonparameteric texture analysis using spatial operators, with applications in visual inspection"*. PhD thesis, University of Oulu, 1997.

[134] G. D. Finlayson and G. Y. Tian. "Colour indexing across illumination". *The proceedings of CIR 99*, 1999.

[135] S. Seitz. "The Space of All Stereo Images". *Proceedings of ICCV 2001*, 1:26–33, 2001.

[136] C. Sun. "Fast Stereo Matching Using Rectangular Sub-regioning and 3D Maximum-surface Techniques". *International Journal of Computer Vision*, 43(1-3):99–117, April-June 2002.

[137] D. Gledhill, G. Y. Tian, D. Taylor, and D. Clarke. "Panoramic imaging - a review". *International Journal of Computers and Graphics*, 27(3):435–445, 2003.