# DESIGN AND IMPLEMENTATION OF A SPATIALLY ENABLED PANORAMIC VIRTUAL REALITY PROTOTYPE

STEPHEN RAWLINSON

January 2002

TECHNICAL REPORT
NO. 215

# DESIGN AND IMPLEMENTATION OF A SPATIALLY ENABLED PANORAMIC VIRTUAL REALITY PROTOTYPE

Stephen Rawlinson

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

February 2002

© Stephen Rawlinson 2002

# PREFACE

This technical report is a reproduction of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Department of Geodesy and Geomatics Engineering, February 2002. The research was supervised by Dr. Y. C. Lee, and it was financially supported by the Natural Sciences and Engineering Research Council of Canada.

As with any copyrighted material, permission to reprint or quote extensively from this report must be received from the author. The citation to this work should appear as follows:

Rawlinson, Stephen (2002). *Design and Implementation of a Spatially Enabled Panoramic Virtual Reality Prototype*. M.Sc.E. thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 215, University of New Brunswick, Fredericton, New Brunswick, Canada, 106 pp.

# ABSTRACT

Conventional approaches to adding virtual reality-based realism in a GIS environment involve the development of complicated 3-dimensional geometric models through the use of sophisticated computer hardware and software. While these approaches provide for some benefits with regard to increased user comprehension, they are often limited due to the complexity of their creation and inability to provide realistic visual cues for the user. This is especially significant in the development of interactive computer-based touring guides, where the uninitiated user must be able to quickly and efficiently interpret directions provided on a computer display.

This research focuses on the integration of digital terrestrial photographs in a map-based environment acquired with a set of non-metric cameras mounted on a simple tripod system. A novel combination of stereo-photographic and image processing techniques are used to link 360-degree panoramic virtual environments to a dynamic map-based environment within a software and hardware prototype developed by the author. The linked panoramic and map interface allows for user query and interaction. Techniques and results are outlined for the creation of the system, including: acquisition, processing (data reduction), and visualization. Ease of use and low cost were primary considerations for the development of the prototype. Results suggest that an un-calibrated stereo and camera setup can provide appropriate accuracy for the purposes of GIS integration.

Further, the successful implementation of the prototype provides proof of concept for an alternative approach for spatially enabled virtual reality and map integration.

# ACKNOWLEDGEMENTS

Many thanks go to my supervisor, Dr. Y.C. Lee, for his continuous ideas and guidance throughout my studies. It is truly an honour to be his first graduate student upon his return to Canada. As well, I would like to thank Dr. David Coleman for his boundless enthusiasm and support for my academic and professional careers. My colleagues, most notably Kevin Pegler, were superb in their heartfelt advice and good humour.

My final thanks, as always, go to my family, Ellen, Alan, Graham, and Geoffrey. Not only have they collectively instilled in me the value of hard work, they have always encouraged to me to excel wherever in the world my career has taken me. This thesis would simply not be possible without their support.

# TABLE OF CONTENTS

**ABSTRACT** ..... **ii**

**ACKNOWLEDGEMENTS** ..... **iv**

**TABLE OF CONTENTS** ..... **v**

**LIST OF TABLES** ..... **viii**

**LIST OF FIGURES** ..... **ix**

**CHAPTER 1 - INTRODUCTION** ..... **1**

1.1 Geometric Modelling ..... 1

1.2 Image-based Rendering ..... 2

1.3 Panoramic Virtual Reality ..... 4

1.4 Current Panoramic VR and GIS Approaches ..... 5

1.5 Current Close-Range/Terrestrial Photogrammetric Approaches ..... 6

1.6 A New Approach for VR GIS ..... 7

**CHAPTER 2 – BACKGROUND** ..... **10**

2.1 Introduction ..... 10

Part I – Photogrammetric Concepts ..... 10

2.2.1 Non-metric close-range photogrammetry ..... 10

2.2.2 A simple stereo system ..... 12

2.2.3 General solution ..... 15

2.2.4 Camera Calibration ..... 20

2.2.5 Collinearity Enforcement ..... 20

2.2.6 Coplanarity Enforcement ..... 22

2.2.7 Camera Calibration Techniques ..... 24

2.2.7.1 Modified Direct Linear Transformation Calibration Technique ..... 26

Part II – Correspondence Problem ..... 28

2.3.1 Image Matching ..... 28

2.3.1.1 Area-based Matching ..... 30

2.3.1.2 Similarity measures ..... 30

2.3.2 Feature-based Matching ..... 34

2.3.3 Epipolar geometry ..... 35

**CHAPTER 3 - METHODS ..... 37**

3.1 System Overview ..... 37

3.2 Image Acquisition ..... 38

3.3 Panoramic Warping ..... 45

3.4 Projection of a plane to a cylinder ..... 47

3.4.1 Resampling Techniques ..... 48

3.4.1.1 Nearest Neighbour Resampling ..... 49

3.4.1.2 Bilinear Interpolation Resampling ..... 50

3.5 Image Alignment ..... 51

3.6 Panoramic Image Sequence Characteristics ..... 52

3.7 Brute Force Correlation Matching ..... 53

3.7.1 Algorithm Description ..... 53

3.7.2 Brute Force Approach Refinements ..... 55

3.8 Adaptive Matching Approach ..... 58

3.8.1 Algorithm Description ..... 60

3.9 Blending and End-to-End Alignment ..... 69

3.10 Visualization ..... 72

3.11 Space Positioning ..... 73

3.11.1 Conversion to Original Image Coordinates ..... 75

3.11.2 Stereo Matching ..... 76

3.11.2.1 Stereo Pair Characteristics ..... 77

3.11.2.2 Algorithm Description ..... 78

3.11.2.3 Algorithm Refinement ..... 79

3.11.3 Calculation of object depth ..... 81

3.11.4 Camera calibration ..... 82

3.11.5 Determination of Bearing ..... 84

**CHAPTER 4 – PROTOTYPE RESULTS AND ACCURACY ASSESSMENT ..... 85**

4.1 Introduction ..... 85

4.2 Stitching (Image Matching) Implementation Evaluation ..... 86

4.3 Distance Calculation Evaluation ..... 90

4.4 System Performance ..... 97

**CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS ..... 99**

**REFERENCES ..... 102**

**VITA ..... 107**

## LIST OF TABLES

**Table 1.1.** Comparison of the geometric modeling and image-based rendering approach to virtual reality (adapted from Kang, 1998). ..... 3

**Table 2.1.** Benefits and limitations of a non-metric camera setup for photogrammetric purposes ..... 11

**Table 4.1.** Translation errors in *x* and *y* for brute force and adaptive matching approaches given an a priori matching estimate of (–180,0), kernel size of 7 x 7 pixels, and 384 x 256 image size ..... 86

**Table 4.2.** Camera calibration results (interior orientation parameters) for the left camera given an image size of 384 x 256 pixels ..... 90

**Table 4.3.** Camera calibration results (interior orientation parameters) for the right camera given an image size of 384 x 256 pixels..... 91

**Table 4.4.** Relative orientation parameters for stereo panoramic image pairs .. 92

LIST OF FIGURES


**Figure 1.1.** Overview of cylindrical panoramic imaging process ..... 4
**Figure 1.2.** Conceptual representation of GIS and panoramic virtual reality integration ..... 8
**Figure 2.1.** Top-down view of a simple stereo system ..... 13
**Figure 2.2.** "Normal Case" configuration for calculation of object space coordinates ..... 15
**Figure 2.3.** Collinearity condition geometry for close-range/terrestrial applications ..... 19
**Figure 2.4.** Coplanarity condition geometry for close-range/terrestrial applications ..... 22
**Figure 3.1.** Spatially enabled panoramic image environment component overview ..... 38
**Figure 3.2.** Overhead view of rotating panoramic and stereo-pair image acquisition ..... 39
**Figure 3.3.** Modified dual panoramic/stereo tripod ..... 41
**Figure 3.4.** Left camera mount showing centre of rotation close to nodal point .. 41
**Figure 3.5.** Right camera mount ..... 41
**Figure 3.6.** Nodal point offset for Olympus TripXB400 camera (bottom view) ... 42
**Figure 3.7.** Left (left column) and right (right column) sequence of camera shots of City Hall, Fredericton, NB, April, 2001 ..... 43
**Figure 3.8.** Population of warped cylindrical matrix using planar matrix ..... 48
**Figure 3.9.** Nearest neighbour versus bilinear interpolation resampling techniques of warped test image (pd = 288 pixels or 27 mm), Fredericton, NB, April, 2001 ..... 49
**Figure 3.10.** Image Distortion as a function of principal distance for image sequence 0, Fredericton, NB, April 2001 (Original image size 384 x 256 pixels) ..... 50
**Figure 3.11.** Constrained search windows for image sequence 0 and 1 showing kernels centred at $(x_0,y_0)$ and $(x_1,y_1)$ respectively, Fredericton, NB, April 2001. ..... 54
**Figure 3.12.** Homogeneity of pixel regions in image matching ..... 57
**Figure 3.13.** Overview of adaptive matching algorithm components for image stitching ..... 59
**Figure 3.14.** Vertical, horizontal, and magnitude outputs for image 0 and 1, City Hall, Fredericton, NB. ..... 61
**Figure 3.15.** Extracted corners using the adaptive corner extraction technique for test image sequence 1, City Hall, Fredericton, NB ..... 66
**Figure 3.16.** Candidate matching support concept ..... 68
**Figure 3.17.** Linear function for blending offset image pairs ..... 70

**Figure 3.18. Seamless output mosaic as the product of automatic warping, alignment, blend, and end-to-end alignment for City Hall, Fredericton, NB, test location ..... 71**
**Figure 3.19. User interface for visualization of panoramic image..... 72**
**Figure 3.20. Stereo matching technique ..... 77**
**Figure 3.21. Two general cases for the refined stereo matching approach..... 80**
**Figure 3.22. Calibration test field developed by Liu, 1991 ..... 82**
**Figure 4.1. Prototype system accuracy: normal and calibrated case ..... 92**
**Figure 4.2. Distribution of 40 manually selected panoramic scene measuring points ..... 93**
**Figure 4.3. Absolute error versus distance for prototype system ..... 94**

# CHAPTER 1 - INTRODUCTION

Virtual reality (VR) is gaining in popularity as a useful visualization technique. VR systems allow for the creation of virtual environments, which place users in a computer simulated environment allowing for interaction (El-Hakim *et al.*, 1998). While VR has it roots in the 1950s and 1960s (Kalawsky, 1993), concerted research into VR systems began only in the 1980s, when computer processor power became sufficiently adequate to allow for effective realism. Since users of VR systems are placed in a computer generated environment, they can be introduced into situations or scenarios that would be unsafe or impractical in the real world. Well known examples of VR systems include airline flight simulators and military battlefield simulators. VR systems have traditionally been developed and designed within the computer graphics community, for example CAVE (Cruz-Neira *et al.*, 1993); and the virtual workbench (Kruger *et al.*, 1995). These systems render 3D geometric models generated from secondary sources such as 3D digitizing tools, rangefinders, and stereo photogrammetric techniques. Surface texture shading or environment maps are subsequently introduced to the models to increase realism (Kang, 1998).

## 1.1 Geometric Modelling

The above VR approach, referred to as geometric modelling (GM), has been adopted by the GIS and cartographic communities and is a growing area of active

research (Germs *et al.*, 1999; Hearnshaw and Unwin, 1994; Huang and Lin, 1999; Rhyne, 1997; Unwin, 1997). Three-dimensional VR GIS is largely focused on the visualization of geographic scenes to mimic human perspective views (Raper *et al.*, 1999). This has traditionally been in concert with the design and development of 3-dimensional topologic models and spatial query techniques (de la Losa and Cervelle, 1999). Interestingly, the first true experimentation with VR GIS began with work to develop efficient data structure translators to move from GIS to VR formats (Raper and McCarthy, 1994). In this regard, a GIS was viewed primarily as a data processing tool, and not a viewing environment.

The popularity of GM VR and GIS can be largely attributed to the decreasing cost and increasing availability of powerful rendering hardware and software, in conjunction with a general awareness in these communities that 3D visualization dramatically increases the level of understanding for the end user. Compared with standard 2D planimetric maps oriented to the north, 3D scenes present almost unlimited viewing perspectives. The availability of commercial GIS software products supporting 3D visualization, such as ESRI's ArcView 3D Analyst extension and ERDAS's VirtualGIS, and the development of a 3D "geographic" modelling language, Virtual Reality Markup Language, or VRML, typify this trend (ESRI, 2001; ERDAS, 2001; VRML, 2001).

### 1.2 Image-based Rendering

Recently however, a new VR approach, called image-based rendering (IBR), has emerged that renders photo-realistic views depending on the user's observation location

(Chen, 1995; Szeliski and Shum, 1995; McMillan, L. and Bishop, G., 1995). Views are represented as a mosaic or collection of images and new views created by interpolating and/or reprojecting input images onto target surfaces such as cylinders, spheres, and more recently cubes (Szeliski and Kang, 1995; APPLE, 2001). As Kang (1998) suggests, this contrasts with the GM approach where the typical rendering process relies on modeling transformation, view transformation, culling (deciding on and displaying what is theoretically visible), and finally hidden surface removal. This is an important difference since increased realism requires increasingly complex geometric models, and thus the cost of rendering in a GM VR can be high since rendering time is a function of the scene complexity. In fact, the GM approach is well known to require "laborious modeling and special purpose software" for effective realistic view rendering (Chen, 1995). This is especially significant when data volumes are high and thus realistic rendering requires high-end PCs with high performance graphics display cards. A comparison of the GM and IVR approaches in the context of GIS is presented in Table 1.

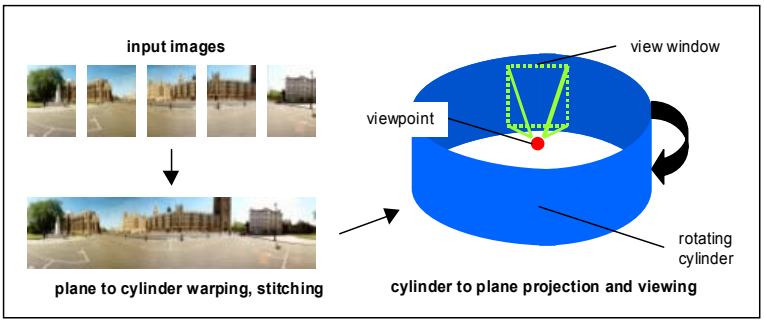
**Table 1.1. Comparison of the geometric modeling and image-based rendering approach to virtual reality (adapted from Kang, 1998).**

| Geometric modeling approach | Image-based rendering approach |
|---|---|
| Complex 3D geometric data structures | Set of images |
| Conventional rendering | Reprojection/Interpolation |
| Sophisticated hardware/software for added realism | Realism function of input scenes |
| Expensive inputs | Inexpensive inputs |
| Query support | Limited query support |
| Link to GIS well developed | Link to GIS less developed |

## 1.3 Panoramic Virtual Reality

Perhaps the most widely known and available IBR technique is panoramic virtual reality, or PVR. This novel VR approach allows for complete 360 degree panning and viewing around a given observation point by warping a set of input images to simulate a user's perspective view. The set of input overlapping images are generally acquired around a rotation point by consecutively panning a camera until complete 360-degree coverage is obtained. These images are subsequently stitched together and warped onto a cylinder to form a continuous mosaic (Figure 1.1). Using a standard desktop PC and appropriate software (such as Apple's QTVR), realistic scenes can be rendered (re-projected from the cylinder onto a plane) "on-the-fly" (Apple, 2001; IPIX, 2001).

**Figure 1.1. Overview of cylindrical panoramic imaging process**




In addition, static "hot-spots" can be created that identify pixel regions on a panoramic image that support additional interaction, such as WWW navigation or activating actions (Chen, 1995). The "hot-spot" concept, while seemingly useful in providing GIS linking capability, are simply user defined pixel regions and thus has no geographically referenced meaning.

### 1.4 Current Panoramic VR and GIS Approaches

Unlike the GM approach, the integration of panoramic VR and GIS is less developed. For example, Chapman and Deacon (1998) used panoramic imagery along with texture mapping to supplement traditional 2D and 3D CAD databases. They manipulated the 2D panoramic image by superimposing a 3D texture mapped CAD model and thus objects could be placed within the scene for a potentially realistic virtual reality representation. However, this approach was limited since a computational approach was not used to establish 3D geometry in the 2D panoramic image. As a result, the illusion of 3D was not consistently manifested since objects were placed manually on an ad-hoc basis.

Furthermore, Dykes (2000) integrated panoramic imaging to a geographic base to provide bearing information in the context of a virtual field course. At each waypoint a panoramic VR scene is created with compass directions being linked to a planimetric map. The approach proved successful in the case of student field research. An extension to the navigation approach is Virtually Vancouver, a commercially based Internet site that provides for an integrated map and panoramic imaging capability at numerous street intersections located in Vancouver, British Columbia, Canada (Virtually Vancouver, 2001).

While these approaches are advantageous over GM techniques due to their simplicity in design and their added realism, they fail to effectively take advantage of the full potential of a dynamic link between a photo-realistic VR environment and a spatial

database. In fact, Chapman and Deacon (1998) suggest that "until we are able to automatically generate 3D geometry from 2D image data it is unlikely to be cost-effective to maintain… databases". The design of a system that generates valuable coordinate information within the PVR environment for linking with a GIS is the focus of this research.

### 1.5 Current Close-Range/Terrestrial Photogrammetric Approaches

While the development of a spatially enabled panoramic viewing system has yet to be developed to the best available knowledge of the author, close-range photogrammetric techniques are well developed in the literature. Liu (1991) developed a stereo system for use in vehicle crash investigations using standard commercially available, and inexpensive, cameras. With moderate camera calibration techniques, he concluded that accurate coordinate information could be calculated given a pair of stereo images. Further, Huang (1998) developed a combination digital photo imaging and theodolite device designed to georeference close-range photographs of building and urban environments suitable for use in a GIS. The system was suitable for use with a land surveyor with little or no extra training.

In addition, there are numerous commercially available close-range photogrammetric software packages on the market, including EOS Systems PhotoModeler Pro™ and Vexcel's FotoG-FMS™. Although there remains some debate about the claims of accuracy and ease of use of these systems, this author suggests that the plethora of desktop PC-based close-range/terrestrial software packages leads to the conclusion that photogrammetric theory has advanced sufficiently for use in diverse

applications. However, to the author's knowledge, there has not yet been a rigorous investigation of the integration of photogrammetric principles within a panoramic imaging environment.

### 1.6 A New Approach for VR GIS

Due to the aforementioned inadequacies of existing panoramic virtual reality and GIS integration approaches, the objective of this research is to develop and test a complete methodology for acquiring, processing, and displaying panoramic images that are linked to a GIS environment. The idea here is to construct a prototype that presents a user with two views of a scene (a standard 2D overhead view and an interactive 360 degree panoramic view) that are dynamically linked such that user interaction in one view is reflected in the corresponding view. Based on this concept, valuable spatially linked attribute information from the GIS can be displayed (Figure 1.2). In addition, the following are key design considerations: 1) low cost; 2) easily available inputs (no object control and little or no calibration); 3) simplicity and ease of use for the non-specialist; 4) adequate accuracy (+/- 1-2 metres) for the purposes of GIS integration and, 5) robustness and reliability. It is anticipated that a full working prototype would find use in interactive touring and navigation guides, as well as planning and view shed visualization.

**Figure 1.2. Conceptual representation of GIS and panoramic virtual reality integration**

The research outlined in this paper presents an alternative approach for GIS and IBR virtual reality integration that provides a true link between the image scene and the GIS database through a prototype georeferenced panoramic imaging environment. This system takes advantage of stereo photogrammetric principles and image processing techniques to provide proof of concept for seamless virtual reality and GIS integration.

Interestingly, the term "photo-spatial" VR has been used, perhaps erroneously, to describe panoramic VR (Dodge *et al.*, 1998). This author suggests that this is misleading and causes confusion for the non-specialist since no 3D coordinate geometry or depth information is implied or computed in standard panoramic VR systems. To the best available knowledge of the author, the approach described herein represents the first attempt to integrate spatial positioning within a panoramic imaging environment. For the purposes of this research, the term "photo-spatial" VR will be used exclusively to denote a VR system explicitly enabling a georeferencing capability.

This thesis is organized into the following sections:

- Chapter 1 (Introduction) provides an introduction into the rationale and scope of this thesis research.
- Chapter 2 (Background) provides 1) a detailed examination and explanation of non-metric close range photogrammetric principles along with an discussion of current non-metric camera calibration techniques; and 2) a overview of current image matching approaches for the purposes of automatic stereo matching and alignment.
- Chapter 3 (Methodology) documents the prototype developed in this research and details each software component developed by the author.
- Chapter 4 (Results and Accuracy Assessment) provides a detailed account of the results obtained, including camera calibration (interior and exterior/relative orientation) parameters, matching accuracy (image stitching), and system accuracy ("normal case" and calibrated setup). The performance of the prototype system is also documented.
- Chapter 5 (Conclusion and Recommendations) discusses the significance of the results obtained and explores further refinements of the prototype system. A summary of the key findings of this research is provided.

# CHAPTER 2 – BACKGROUND

## 2.1 Introduction

It is a well-known photogrammetric principle that if an object appears in two or more images acquired from at least two sensor locations, the spatial position of the object can be determined by intersecting collinear rays. This principle, referred to as space intersection, can be used to calculate the 3-dimensional position of an object contained within at least two 2-dimensional photographs.

From a computational perspective, even the most basic stereo system must solve two fundamental problems. The first problem relates to the geometry of the cameras internally and with respect to each other. This is a standard photogrammetric problem and a description of available solutions found in the literature forms the first part of this chapter. The second problem consists of determining which feature in the left camera corresponds to which feature in the right camera. An automated approach for solving this second fundamental problem is non-trivial and a detailed discussion forms the second part of this chapter.

## Part I – Photogrammetric Concepts

### 2.2.1 Non-metric close-range photogrammetry

The conventional approach for distance and space positioning commonly used in aerial photogrammetry involves the use of a metric sensor, whose internal geometry is experimentally known and stable. In contrast, the terrestrial approach used in this research relies on the use of a non-metric sensor, whose internal geometry is not known

and not always stable. As a result, the standard photogrammetric data reduction and data evaluation procedures used for metric cameras are not appropriate for non-metric imageries (Faig, 1989). More sophisticated processing procedures are generally required for non-metric imagery, especially if the same or nearly the same accuracy is desired. Non-metric cameras are advantageous over metric cameras for the purposes of this research due to their low weight, small size, low cost, and generally wide availability.

The term "close-range photogrammetry" is generally reserved for applications dealing with photographs taken with cameras located on the surface having object distances up to approximately 300 metres (Wolf, 1983; ASPRS, 1980). Close-range photogrammetric image acquisition with non-metric camera systems is becoming more and more popular in engineering and computer vision disciplines. This is largely a result of the evolution of photogrammetric theory as well as the development of more efficient computer processing hardware and software. In this sense, non-metric data reduction and evaluation has been practical for the non-specialist only recently. Further, a growing body of scientific literature has supported the notion that high levels of accuracy can be obtained from non-metric inputs. Table 2.1 presents the advantages and disadvantages of the more general case of non-metric imaging systems over traditional metric system.

**Table 2.1. Benefits and limitations of a non-metric camera setup for photogrammetric purposes**

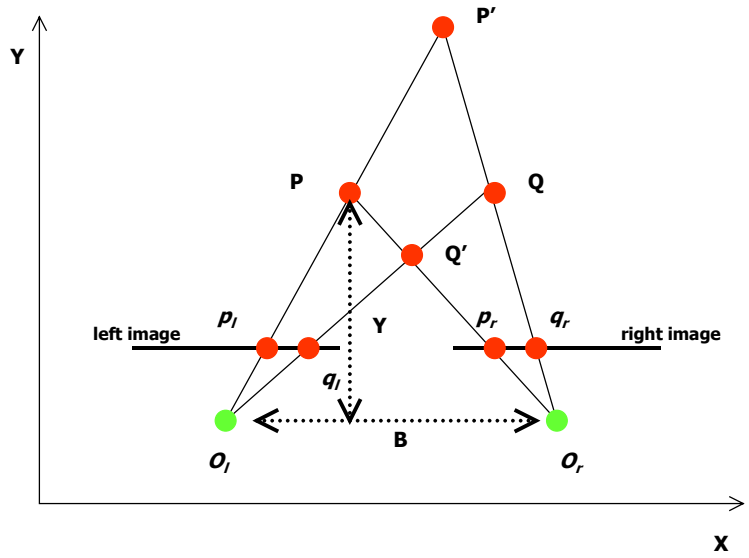
| Advantages | Disadvantages |
|---|---|
| ▪ General availability<br>▪ Portable, light-weight<br>▪ Low cost | ▪ High distortion (lens)<br>▪ Unstable interior orientation<br>▪ Lack of fiducial marks<br>▪ Non-trivial determination of exterior orientation during acquisition |

Although it is clear that non-metric camera systems differ markedly from metric camera system, both systems basically function as central perspective imaging systems. As such, both rely on the principle of collinearity; that is, that image point, exposure centre, and object point all lie along a straight line. This assumption is critical as it serves as the foundation for analytical photogrammetric calculations and is used extensively to establish the relationship between image and object space coordinates. The collinearity principle is outlined in the next section.

### 2.2.2 A simple stereo system

The computation of object positions in space requires linking coordinates of points in 3-dimensional space with coordinates of their corresponding images points (and vice-versa). Using the example of a simple stereo system (Figure 2.1) shown from a top-down perspective, the determination of the position of $P$ and $Q$ in space can easily be seen through the process of triangulation. Triangulation, in turn, depends crucially on the solution of the correspondence problem; that is, the calculation of $(p_l, p_r)$ and $(q_l, q_r)$. The correspondence problem is discussed in detail in part II of this chapter; thus, for the purposes of this discussion, consider the correspondence problem solved.

**Figure 2.1. Top-down view of a simple stereo system highlighting significance of correspondence problem**




Thus, the intersection of rays $O_lp_l - O_rp_r$ and $O_lq_l - O_rq_r$ leads to interpreting the image points as projections of $P$ and $Q$. However, if $(p_l,q_r)$ and $(q_l, p_r)$ are the selected pairs of corresponding points, triangulation returns $P'$ and $Q'$. Note that both interpretations, although radically different, are equally valid.

Figure 2.1 also illustrates the triangulation of a single point $P$, determined from image coordinates $p_l$ and $p_r$. The distance, $B$, between the centres of projection $O_l$ and $O_r$, is referred to as the baseline of the stereo system. Letting $x_l$, $x_r$ be coordinates of $p_l$, $p_r$ with respect to the principal points $c_l$, $c_r$, $f$ the camera constant or focal length, and $Y$ the distance between $P$ and the baseline. From the similar triangles $(p_l, P, p_r)$ and $(O_l, P, O_r)$, the following equation can be derived:

$$\frac{B + x_l - x_r}{Y - f} = \frac{B}{Z} \qquad (2.1)$$

Solving for $Y$ yields the following:

$$Y = f\frac{B}{d} \qquad (2.2)$$

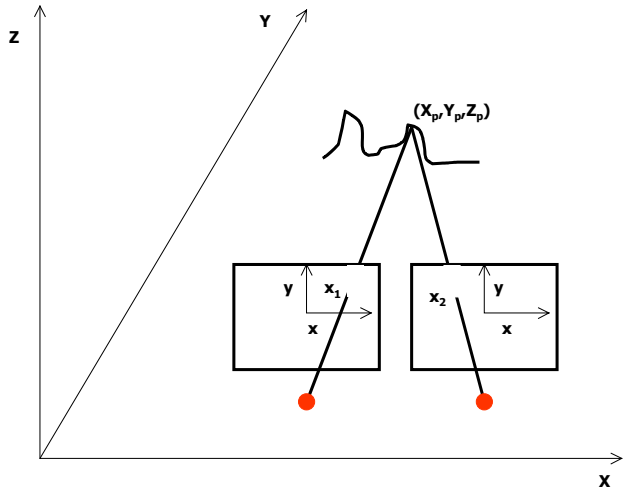
where $d = x_r - x_l$ and is referred to as the disparity or parallax. Parallax measures the difference in retinal position between the corresponding points in the two images. As such, depth is inversely proportional to parallax. Equation 2.1 is the simple case (normal case) and although it is computationally attractive, it is rarely encountered in real world applications. In a 3-dimesional perspective and with the normal case assumption, any image object space ($X_p$, $Y_p$, $Z_p$) 3-dimensional coordinate corresponding to a feature contained within the left and right images can be calculated using the following sets of equations (Figure 2.2)

$$X_p = \frac{Yp}{pd} x_1$$
$$Y_p = \frac{Bpd}{x_1 - x_2}$$
$$Z_p = \frac{Yp}{pd} y \qquad (2.3)$$

where :

$x_1$ = $x$ position of the object in the left image (panorama)
$x_2$ = $x$ position of the object in the right image
$pd$ = estimated principal distance
$B$ = stereo base separation

**Figure 2.2. "Normal Case" configuration for calculation of object space coordinates**




### 2.2.3 General solution

Two concepts complicate the geometry of Figures 2.1 and 2.2; namely, the internal geometry of the camera and image deviates from the ideal central perspective camera model, and corresponding exposure centres of the left and right cameras may not always be perpendicular to the stereo system baseline.

The internal geometry of the camera is characterized by a set of interior orientation parameters: principal distance $f$ (perpendicular distance from the perspective centre of the lens to the image plane), principal point $(x_c, y_c)$, and a set of distortion parameters introduced by the optics of the camera. There are many mathematical models found in the literature to date for characterizing lens distortion; however, it is generally accepted in the photogrammetry community that radial and tangential distortions can effectively characterize the majority of lens distortions commonly encountered. In general, knowledge of the interior orientation parameters is sufficient to describe the

optical, geometric (relationship between the perspective centre and the image plane), and digital characteristics of the viewing camera.

An additional interior orientation parameter is the transformation between the image frame coordinates and pixel coordinates. A digital image is made up of a rectangular matrix of pixels, each with a specific digital numerical value or set of values (as is the case with colour imagery). As is customary in image processing, the origin of the image coordinate system is in the upper left corner of the image matrix. In this way, the basic unit of the image coordinates is pixels, and thus a conversion between image pixels and metric units is possible as follows:

$$\begin{aligned} x &= -(x_{image} - x_c)s_x \\ y &= -(y_{image} - y_c)s_y \end{aligned} \qquad (2.4)$$

where $(x, y)$ is the image coordinate expressed in metric units, $(x_{image}, y_{image})$ is an arbitrary image point in pixel units, $(x_c, y_c)$ is the principal points in pixel units, and $(s_x, s_y)$ is the effective pixel size in the horizontal and vertical directions respectively (typically expressed in millimetres).

Similarly, the camera reference frame is often unknown, and exterior orientation parameters are defined as any set of geometric parameters that uniquely identify the transformation between the unknown camera reference frame and a known reference frame. In general, the transformation between two frames can be described by a 3-dimensional translation vector $\mathbf{T}$ (where $\mathbf{T} = [x_o\ y_o\ z_o]^T$ ), outlining the relative positions

of the origins of the two reference frames; and a 3 x 3 rotation matrix **R**, an orthogonal matrix ($\mathbf{R}^{\mathsf{T}}\mathbf{R} = \mathbf{R}\mathbf{R}^{\mathsf{T}} = \mathbf{I}$) that brings the corresponding axes of the two frames onto each other. Each camera in the stereo system can be described by the six exterior orientation elements: a 3-dimesional coordinate of the perspective centre ($X_{ol}$,$Y_{ol}$,$Z_{ol}$) and three orientation angles. The rotation is represented using Euler angles ω, φ, κ that define a sequence of three elementary rotations around *x, y, z* axis respectively. The rotations are performed clockwise, first around the *x*-axis, then the *y*-axis that is already once rotated, and finally around the *z*-axis that is twice rotated during the previous stages.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathrm{R}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathrm{T}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \tag{2.5}$$

$$\begin{aligned} m_{11} &= \cos\varphi\cos\kappa \\ m_{12} &= \sin\varpi\sin\varphi\cos\kappa - \cos\varpi\sin\kappa \\ m_{13} &= \cos\varpi\sin\varphi\cos\kappa + \sin\varpi\sin\kappa \\ m_{21} &= \cos\varphi\sin\kappa \\ m_{22} &= \sin\varpi\sin\varphi\sin\kappa + \cos\varpi\cos\kappa \\ m_{23} &= \cos\varpi\sin\varphi\sin\kappa - \sin\varpi\cos\kappa \\ m_{31} &= -\sin\varphi \\ m_{32} &= \sin\varpi\cos\varphi \\ m_{33} &= \cos\varpi\cos\varphi \end{aligned}$$

where $[x\ y\ z]^{\mathsf{T}}$ are coordinates expressed in the camera frame, the *m*'s denote elements in the rotation matrix **R** above, $[X\ Y\ Z]^{\mathsf{T}}$ is the location of an arbitrary point in space.

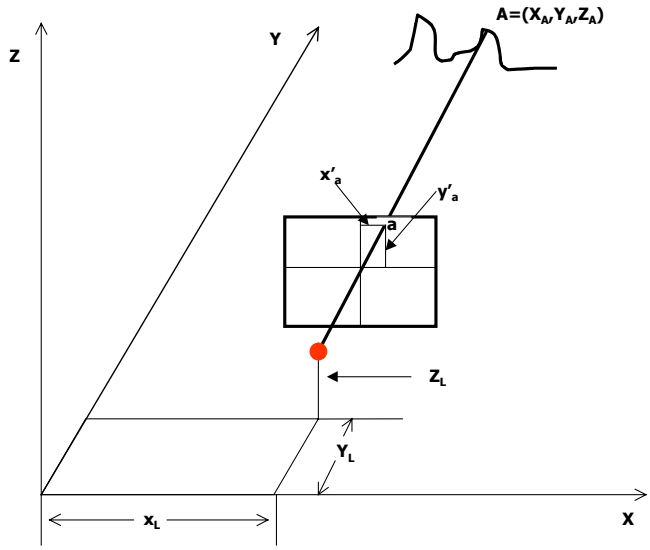
The "known" reference frame is generally considered the world reference frame (absolute orientation); however, for the purposes of a stereo system, the reference frame of one camera (either the left or right camera) can be considered the world reference frame (relative orientation). This assumption is sufficient for the reconstruction of depth information from a set of stereo images.

A general solution to the problem of relating image and object space coordinates in terrestrial or close-range photogrammetry is illustrated in Equation 2.6.

$$\begin{aligned} x_a &= -f \frac{m_{11}(X_A - X_L) + m_{12}(Z_A - Z_L) + m_{13}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)} \\ y_a &= -f \frac{m_{21}(X_A - X_L) + m_{22}(Z_A - Z_L) + m_{23}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)} \end{aligned} \qquad (2.6)$$

These equations are known as the collinearity equations, where the *m*'s denote the elements of the rotation matrix **R** described above, *f* refers to the camera principal distance, $x_a$, $y_a$ are some arbitrary image coordinates, $(X_L, Y_L, Z_L)$ is the 3-D location of the camera in a given reference frame, and $(X_A, Y_A, Z_A)$ is the 3-D location of the object *A* in space. The geometry of the terrestrial/close-range collinearity condition is shown in Figure 2.3.

**Figure 2.3. Collinearity condition geometry for close-range/terrestrial applications**




The collinearity condition can be extended to a stereo system by setting up two sets of equations for both the left and right exposure station for a total of 25 variables (3 elements of the interior orientation, 6 elements of the exterior orientation, 2 coordinates of the projected point in each image, and 3 object space coordinates). The object space location of any given point, for example $(X_A, Y_A, Z_A)$, can be determined by solving all four equations through a least squares adjustment if all variables except the 3 object space coordinates are assumed known.

In the case of metric camera systems, appropriate corrections to account for such phenomena as lens distortion can be applied to the image coordinates prior to the adjustment. This process, known as image refinement, is not effective in the case of non-metric camera systems due to their instabilities and thus image refinement for non-metric imageries are not effective for optimal accuracy (Faig, 1989). In addition, non-metric cameras are not designed for photogrammetric applications; as such, the interior

orientation parameters are unknown. Calibrating the camera yields a set of orientation parameters and is the focus of the next section.

### 2.2.4 Camera Calibration

Camera calibration is usually carried out using a physical or an analytical approach. The physical approach relies on optical equipment in a laboratory setting to determine the physical properties of the camera under consideration. This is the domain of a metric camera system, and thus is not considered further in this thesis.

The analytical approach enforces specific geometric conditions, such as collinearity or coplanarity (or both), that require object space control and no object space control respectively. While the analytical approach is used extensively for metric cameras, it also finds relevance for non-metric cameras as well.

### 2.2.5 Collinearity Enforcement

The condition expressed by the collineation of three points: object point, image point, and perspective centre, expressed in (2.6) and referred to as the collinearity equation, can be extended to model a distorted central projection through the introduction of additional parameters (equation 2.7). In general, no absolute collinearity condition exists due to the physical limitations of the camera and external factors, such as the atmosphere. However, by introducing additional parameters, an optimal fit to the collinearity condition can be readily achieved.

$$x_a + \Delta x_p = -f \frac{m_{11}(X_A - X_L) + m_{12}(Z_A - Z_L) + m_{13}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)}$$
$$y_a + \Delta y_p = -f \frac{m_{21}(X_A - X_L) + m_{22}(Z_A - Z_L) + m_{23}(Y_L - Y_A)}{m_{31}(X_A - X_L) + m_{32}(Z_A - Z_L) + m_{33}(Y_L - Y_A)} \quad (2.7)$$

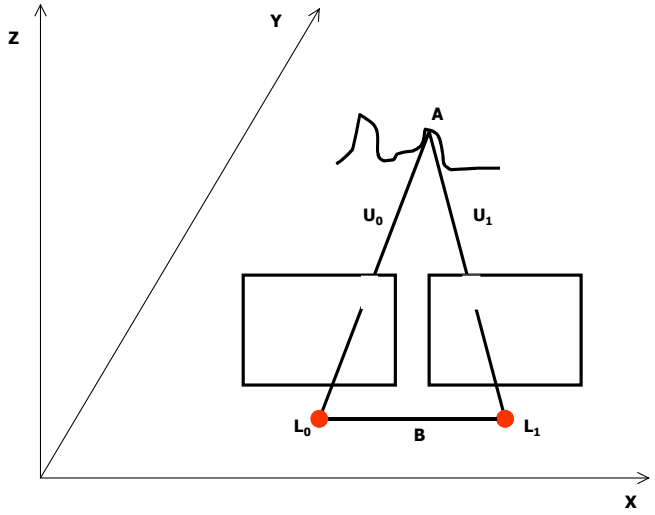
where $x_p$ and $y_p$ represent functions of several unknown additional parameters. These are subsequently adjusted simultaneously with the other unknowns for a complete solution. In all, as long as there are known object space control points, the interior, exterior, and additional parameters can be computed. The number of object space control points required is a function of the number of additional parameters being taken into account. As suggested by Liu (1991), the addition of redundant control points is useful for assessing the calibration accuracy. This technique is often referred to as the single photo calibration approach.

Additional parameter introduction permits the modelling of a non-metric camera and is one of the most commonly used approaches in camera calibration, such as UNBASC2 developed at the University of New Brunswick, Canada. Many different sets of additional parameters have been developed over the past decade, each with its own limitations and benefits. Two categories of additional parameter functions dominate the literature: physical models for modelling the causes or algebraic models for modelling the effects of lens distortions. Interestingly, although the latter generates superior geometric conditions (due to the fact that the additional parameters are usually uncorrelated), superior results are not guaranteed.

### 2.2.6 Coplanarity Enforcement

Coplanarity expresses the relationship between a set of overlapping stereo images that sets the condition that two exposure stations, any object point, and its corresponding image points on the two images all lie in the same plane (Figure 2.4).

**Figure 2.4. Coplanarity condition geometry for close-range/terrestrial applications**




The above figure exists where **B** is the base vector from perspective centres $L_0$ and $L_1$, $\mathbf{U_0}$ is the vector formed by the left perspective centre $L_0$ and some arbitrary point $A$, and $\mathbf{U_1}$ is the vector formed by the right perspective centre $L_1$ and the same arbitrary point $A$. Thus, **B**, $\mathbf{L_0}$, and $\mathbf{L_1}$ all exist within the same plane (ie. they are coplanar). Mathematically, the coplanarity condition can be expressed as:

$$\begin{vmatrix} B_x & B_y & B_z \\ u_0 & v_0 & w_0 \\ u_1 & v_1 & w_1 \end{vmatrix} = 0 \qquad (2.8)$$

*where*

$$\begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix} = U_0 = R_0^{\mathrm{T}} \begin{pmatrix} x_{a0} \\ y_{a0} \\ -f \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}$$

$$\begin{pmatrix} u_1 \\ v_1 \\ w_1 \end{pmatrix} = U_1 = R_1^{\mathrm{T}} \begin{pmatrix} x_{a1} \\ y_{a1} \\ -f \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ w_1 \end{pmatrix}$$

where ($a_{x0}$, $a_{y0}$) and ($a_{x1}$, $a_{y1}$) are image space coordinates of an arbitrary point $A$, $\boldsymbol{R_0}$ and $\boldsymbol{R_1}$ are rotation matrices, and $f$ is the focal length of the camera. It should be noted that $f$ might not be identical for each exposure station in a non-metric stereo system.

In a similar way to that of collinearity, additional parameters can be introduced to account for distortions in the "ideal" coplanar geometry. In contrast to the collinearity approach, camera calibration using the coplanarity condition does not require object space 3D coordinate information. The coplanarity condition is useful for the determination of the relative orientation of the two cameras in the stereo system. In the dependent pair relative orientation approach, the origin of the world coordinate system is assumed to be camera $L_0$, the stereo baseline is assumed to have some unit length in the $x$-component; therefore, this reduces the coplanarity equation as follows:

$$\begin{vmatrix} 1 & b_y & b_z \\ a_{x0} & a_{y0} & -f \\ u_1 & v_1 & w_1 \end{vmatrix} = 0 \qquad (2.9)$$

Thus, only five independent entities are unknown ($b_y$, $b_z$, and the Euler rotation angles ω, φ, κ of the right camera station). Essentially, the dependent approach shifts ($b_y$ and $b_z$ are ratios of the known quantity $b_x$, which is the stereo baseline separation) and rotates the right camera position onto the fixed left camera position.

### 2.2.7 Camera Calibration Techniques

Several techniques for geometric camera calibration can be found in the computer vision and photogrammetry literature. There are three basic approaches that enforce[1] collinearity, coplanarity or both: pre-calibration, on-the-job calibration, and self-calibration.

***Pre-calibration:*** Pre-calibration is the traditional approach for camera calibration. As the name suggests, pre-calibration involves the determination of the camera calibration parameters that are subsequently considered as known entities in further processing (such as space intersection). While pre-calibration does include laboratory methods, it also includes field based methods, where well identifiable and known targets are placed throughout the camera's field of view and subsequently evaluated and processed to form a solution by enforcing one of the geometric conditions. In the case of a non-metric camera setup, it is essential to fix the focus setting on the camera so as to preserve the accuracy of the calibrated parameters.

---

[1]Interestingly, there are novel camera calibration techniques that do not explicitly take into account collinearity or coplanarity. These are not considered further in this report.

***On-the-job calibration:*** On-the-job calibration differs from pre-calibration in that all input imagery serves dual purposes: calibration and evaluation. This means that object space control must exist around the object of interest in order to obtain a solution.

***Self-calibration:*** This technique requires no object space control since it uses the geometric strength of overlapping photographs to determine the parameters of interior orientation plus distortion together with the object evaluation (Faig, 1989). This technique has found applicability in non-metric data reduction schemes; however, it does require sophisticated software processing modules and is computationally expensive.

There are many different techniques presented in the literature for calibrating a non-metric camera system. Each varies in its level of computational efficiency, ease-of-use, availability, and requirement of object space control. The presentation here is not meant to be an exhaustive examination of all calibration techniques or a validation of any particular camera calibration approach; as such, for a complete overview of calibration approaches and techniques, the interested reader is referred to Faig (1989) and Faugeras *et al.* (1992). For the purposes of this research, there are two assumptions that dictated the selection of one particular camera calibration technique:

1. A test field with 61 3D geodetically controlled control points that was used previously by Liu (1991) for camera calibration studies was available for use;

2. A camera calibration software module was made available to the author for unlimited use and had, in previous research, proven to provide acceptable camera calibration accuracy.

#### 2.2.7.1 Modified Direct Linear Transformation Calibration Technique

The Direct Linear Transformation (DLT) calibration approach was originally developed by Abdel-Aziz and Karara (1971). It is based on a modified collinearity condition and can be solved in a closed-form linear fashion. This allows for maximum computational efficiency. The DLT technique has been modified by Heikkila and Silven (1997) to include both radial and tangential image distortion components. Interestingly, Karara and Abdel-Aziz (1974) also included both radial and tangential distortion effects in the formulation of a modified DLT method.

It has been accepted theoretically and confirmed experimentally that a lens system presents two major distortion characteristics, namely radial and tangential (Moniwa, 1980). Radial distortion displaces image points radially outwards or inwards from the optical axis, whereas tangential distortion results from the imperfection of the centering lens in a camera such that nodal point connection (in a compound lens system) is not straight. Radial and tangential distortion is expressed in the modified DLT approach as follows:

$$\begin{aligned} x' &= x(1 + K_1 r^2 + K_2 r^4) + 2P_1 xy + P_2(r^2 + 2x^2) \\ y' &= y(1 + K_1 r^2 + K_2 r^4) + P_1(r^2 + 2y^2) + 2P_2 xy \end{aligned} \tag{2.9}$$

where $x'$, $y'$ are corrected image coordinates, $x$, $y$ are the distorted image coordinates, $r^2=x^2+y^2$, $K_1$ and $K_2$ are the coefficients of radial distortion, and $P_1$ and $P_2$ are coefficients of tangential distortion. As such, the DLT approach reduces to the following equation:

$$\begin{aligned} x(1+K_1r^2+K_2r^4)+2P_1xy+P_2(r^2+2x^2) &= \frac{L_1X+L_2Y+L_3Z+L_4}{L_9X+L_{10}Y+L_{11}Z+1} \\ y(1+K_1r^2+K_2r^4)+P_1(r^2+2y^2)+2P_2xy &= \frac{L_5X+L_6Y+L_7Z+L_8}{L_9X+L_{10}+L_{11}Z+1} \end{aligned} \quad (2.10)$$

where $L_1 \ldots L_{11}$ represent unknown transformation coefficients and the other parameters are as above. Given that each 3D control point ($X,Y,Z$) corresponds to two equations and there are a total of 15 unknowns, a minimum of 8 3D control points are required for a solution. As suggested by Heikkila and Silven (1997), the parameters $L_1 \ldots L_{11}$ do not have any physical meaning; thus, they used an approach developed by Melen (1994) to extract a set of interior orientation parameters from the DLT coefficients. Knowledge of the interior orientation parameters for a given camera permits the determination of the exterior orientation parameters using the dependent pair relative orientation approach as discussed above.

This modified DLT approach is a combination of the pre- and on-the-job calibration approach presented earlier in this thesis. This modified DLT approach is currently implemented in a Matlab programming module and has been made available to this author for use in this research.

## Part II – Correspondence Problem

### 2.3.1 Image Matching

Whether the camera is metric or non-metric, image coordinates of an object from overlapping images play a key role as inputs for space intersection calculations. The quick and accurate determination of these image coordinates through computer automated procedures is therefore of paramount significance. This process is commonly referred to as image matching, while in the more specific case of stereo imagery it is referred to as stereo matching or correspondence matching. A detailed examination of the stereo matching problem and current approaches for finding is the focus on this section.

Interestingly, matching images of the same scene remains one of the most difficult to solve bottlenecks in computer vision (Zhang *et al.*, 1995). At its most rudimentary level, the correspondence problem depends on two key assumptions about the nature of the images (consider two images for the sake of simplicity) under consideration:

1. Most scene points are visible from both camera stations
2. Corresponding image regions are similar

Clearly, both assumptions in real life situations may not hold. In this case, the solution to the correspondence problem is exceeding difficult if not impossible[2]. As Heike (1997) suggests:

> *"Image matching is an ill-posed problem for various reasons. For instance, for a given point in one image, a corresponding point may not exist due to occlusion, there may be more than one possible match due to repetitive patterns or a semi-transparent object surface, and the solution many be unstable with respect to noise due to poor texture"*

However, if it is accepted that both assumptions are valid, then the correspondence problem reduces to a search problem; that is, given a feature in one image, find the corresponding feature in the other image. From a computational perspective, this search problem focuses on two key decisions: 1) which image features should be matched up; and 2) which measure of similarity should be used. It is worthy to note that in the case of stereo matching, condition 1 is highly restricted; that is, the luxury of selecting good features to match up may not be possible given that the user is often tasked with providing this initial location.

Although, perhaps not surprisingly, there are many different approaches for addressing the above, the two fundamental categories are area and feature based. Area based algorithms operate globally on the image pixels, while feature based techniques rely on extracting a subset of pixel locations with which to match. Conceptually, both techniques are essentially indistinguishable; however, their implementations are rather unique. For the purposes of this research, the concept of image matching for photogrammetric stereo applications is considered a subset of the more general image

[2] Clearly, the assumption that both images correspond to the same scene may not hold in real world applications.

matching problem. While stereo matching does in fact inherit many of the assumptions, constraints, and possible solutions developed for image matching in general, it also possesses its own unique constraints. These constraints will be highlighted throughout the discussion below.

#### 2.3.1.1 Area-based Matching

An area based implementation proceeds as follows: for each location in one image (target), find the corresponding location in the other image (search) that maximizes some kind of matching quality measure. Generally, a template window is shifted pixel by pixel across a larger search window, and in each position a similarity between the target template window and the corresponding region of the search window is computed. The optimum value of the similarity measure defines the position of best match between the template and the search window and thus the most likely match.

#### 2.3.1.2 Similarity measures

Various types of similarity measures have been documented for the purposes of image matching. The most common techniques involves the computation of a normalized cross-correlation statistic at each target and search location between the two images. A mathematical expression describing the cross correlation approach is as follows:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}\sigma_{yy}}}, -1 \le \rho \le 1$$

$$\bar{x} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{j} x_{ij}, \qquad \bar{y} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{j} y_{ij} \tag{2.11}$$

$$\sigma_{xx} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{j} x^2{}_{ij} - \bar{x}^2, \quad \sigma_{xy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{j} x_{ij}y_{ij} - \bar{\bar{x}}\bar{y}, \quad \sigma_{yy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{j} y^2{}_{ij} - \bar{y}^2$$

where $p$ denotes the normalized cross-correlation coefficient between the target ($x$) and source ($y$) image, $i = j$= row and column number index in the kernel, and $n$ is the size of the window kernel. An optimal match manifests a normalized cross-correlation value of 1 since, in theory, the stronger the correlation between the two response windows, the more likely that they correspond to homologous features or objects within the images.

Another approach, termed here under the general category of colour separation techniques, involve the comparison of two images based on their absolute intensity level differences. Colour separation techniques have evolved due to the fact that correlation approaches were originally designed for monochromatic (ie. black and white) images. As such, the absolute encoded RGB intensity information corresponding to a true colour image are not used effectively. Furthermore, correlation matching explicitly uses only intensity similarity which can be a weak constraint given that the same intensity value may correspond to a wide array of different RGB colours. A measure of colour separation can be expressed as:

$$C_{sep} = C_R + C_G + C_B \tag{2.12}$$

where

$$C_R = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{j}\left\|R_l(ij) - R_r(ij)\right\|}{256n^2}$$

$$C_G = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{j}\left\|G_l(ij) - G_r(ij)\right\|}{256n^2} \qquad (2.13)$$

$$C_B = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{j}\left\|B_l(ij) - B_r(ij)\right\|}{256n^2}$$

where $C_R$, $C_G$, $C_B$ correspond to the red, green, blue similarity components of the image, and the subscripts *l* and *r* refer to the left and right images respectively, $C_{sep}$ denotes the colour separation measure, and *n* refers to the size of the kernel used to calculate the similarity measure. It should be pointed out that the numerical value "256" in the equation above denotes the dynamic range of the each of the RGB colour components; that is, an RGB image is made up of 8-bit values of red, green, and blue for a complete 24-bit colour range.

Similar to the normalized cross-correlation measure, the colour separation measure denotes the most likely match when it is at its maximum. Interestingly, it has been shown that the similarity components $C_R$, $C_G$, $C_B$ may not always obtain maximum values at the same location due primarily to image noise and contrasting scene illumination conditions (El-Ansari *et al*., 2000). The colour separation measure, $C_{sep}$, can be classified into one of the four proceeding categories:

1. $C_R$, $C_G$, $C_B$ all exhibit maximum values at the same location
2. Only two of $C_R$, $C_G$, $C_B$ exhibit maximum values at the same location

3. Only one of $C_R$, $C_G$, $C_B$ exhibit maximum values at the same location
4. None of $C_R$, $C_G$, $C_B$ exhibit maximum values at the same location

It is perhaps not surprising that the optimal match location occurs when the condition 1 is achieved. There are many deviations from these two similarity matching approaches presented here; however, for the purposes of this thesis, the correlation and colour separation based techniques form the fundamental approaches found in the literature.

As noted by Heike (1997), Gilles (1996), and El-Ansari *et al.* (2000) a central problem with area based techniques is to find the optimal size of the target and search templates. If the region is too small, a wrong match might be found due to ambiguities and noise. If the region is too large, it can no longer be matched as a whole due to occlusions and differences in viewpoint geometry. Further, colour separation techniques, while theoretically attractive, fail to function effectively in a real world scenario where scene variability can be high. While techniques such as histogram matching can be employed to reduce this type of radiometric variability, it is the experience of the author that this type of solution provides very limited benefit at best.

Unfortunately, there is no accepted standard for deciding upon appropriate template sizes in either the computer vision or photogrammetric fields. As a result, while the implementation of area-based techniques can be far easier than that of the feature based approaches due to their reduced complexity, the process of achieving good matching results can be a matter of trial and error and experimentation. Despite this

limitation, area based approaches (including the various derivatives thereof) are arguably the most often used technique used at present in image matching, from remote sensing and photogrammetric applications (Drewnoik and Rohr, 1996; Malmstrom, 1986) to large scale robotic and computer vision applications (Anandan, 1989; Woodfill and Zabih, 1991).

### 2.3.2 Feature-based Matching

In contrast with area based approaches, feature based image matching techniques generally involve the extraction of distinctive features (such as edges) before the matching process, thereby attempting to decide "off-line" which locations in the image should be matched with relatively high confidence. In this way, all other image features are ignored in the subsequent matching process. Typically, the measure of similarity is based on a characteristic of the entity being extracted; for example, if edge information is extracted, then orientation and length could be used as a measure of similarity.

Feature based techniques can prove to be faster than conventional area based approaches; however, it is worthy to note that any specific examination of computational cost must include the cost of producing the initial feature descriptors. Further, feature based techniques can be advantageous over area based approaches since, in general, they are relatively insensitive to illumination changes between corresponding scenes. There are numerous examples of feature-based approaches found in the literature, including Grimson (1985), Huttenlocher *et al.* (1983), Koller *et al.* (1993), and Zhang *et al.* (1995).

Recent advances in feature-based techniques involve the use of dynamic programming to select the optimal match. In its most basic form, dynamic programming focuses on the optimization of a problem in a multi-stage decision process, whereby rules are established to arrive at an incremental decision to the complex problem of stereo matching. Perhaps not unexpectedly, this type of image matching approach can be complex to implement and even more complex to execute in a real world situation. The interested reader is referred to Bernard *et al.* (1986) and Chai and De Ma (1997) for a more complete discussion; this approach is not considered further in this thesis due to the complexity in its design and relatively high computational cost. This makes this approach less suitable in this research application.

### 2.3.3 Epipolar geometry

While the specific case of stereo matching shares all of the characteristics of the general image matching problem discussed above, there is one geometric constraint, namely, the epipolar constraint, that can be exploited to provide a more computationally attractive solution to the stereo matching problem.

The epipolar constraint can be understood through a re-examination of the coplanarity condition expressed in Figure 2.4. In this illustration, $\mathbf{U_0}$, $\mathbf{A}$, and $\mathbf{U_1}$ all define the same plane. This plane can be referred to as the epipolar plane. The image of the projection centre of one camera in the other defines the epipole. Furthermore, the epipolar line is defined as the intersection of the epipolar line and the image plane. With the exception of the epipole, only one epipolar line goes through any image point. The

fundamental concept worth noting is that the epipolar constraint establishes the mapping between points in the left image and lines in the right image, and vice-versa. As such, corresponding points must lie on the on the conjugate epipolar line. Therefore, in order to determine the mapping between points on the left image and corresponding epipolar lines in the right image, the search for a match of the point from the left image can be restricted to those pixels along the corresponding epipolar line. Thus, the search for a match reduces from a 2-dimensional matching problem to a more manageable 1-dimensional problem.

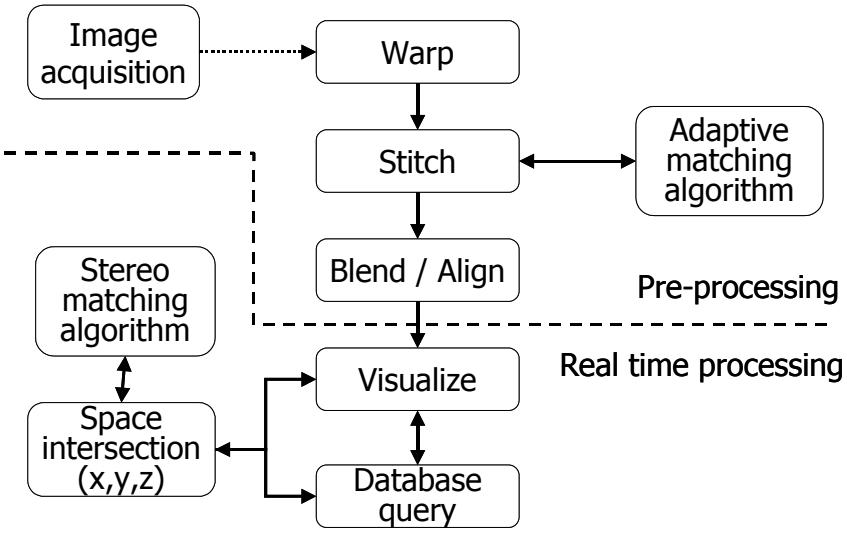
Although enforcing the epipolar constraint may appear to be beneficial, it should be pointed out that determining the appropriate translation and rotation relating the two images and thus establishing the epipolar geometry requires, in practice, another set of corresponding matching locations. Thus, considerable effort and time resources are still required. This author suggests that establishing the epipolar constraint simply shifts the majority of processing from the stereo matching step to the image matching step.

# CHAPTER 3 - METHODS

## 3.1 System Overview

This chapter focuses on the physical design of the prototype developed in this research. The prototype system outlined in this research thesis consists of a conventional survey quality tripod, a modified prototype stereo bar, two (2) identical off-the-shelf cameras (conventional Olympus TripXB400 fixed focus camera, nominal focal length = 27mm, shutter speed 1/100, 24 x 36 mm image size), and a series of software modules. Sets of stereo-pairs corresponding to a complete 360-degree rotation around a desired viewpoint are first acquired using a tripod mount. The images are entered as input into software developed by the author that 1) warps and stitches the imagery into a cylindrical panoramic mosaic; 2) processes the stereo-pairs for further distance calculations (through space intersection); and, 3) displays and renders the mosaicked imagery into the integrated panorama and GIS system for subsequent user query and interaction (Figure 3.1). The linking of the GIS and the panoramic viewer is accomplished internally through the automatic calculation of depth information (distance from the camera to the object under consideration) from the input stereo-pairs. At the time of writing, the entire process is semi-automated and requires very little direct human interaction.

**Figure 3.1. Spatially enabled panoramic image environment component overview (note: each component corresponds to a software module developed by the author)**

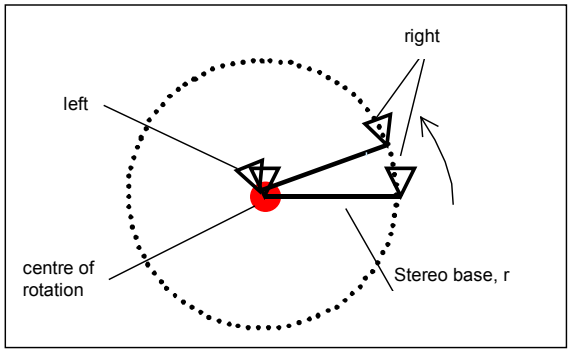



## 3.2 Image Acquisition

The prototype system developed in this research is based on the cylindrical panorama model due to the simplicity in image acquisition and the relative ease in projecting from the cylinder to the plane (and vice-versa).

Stereo-pairs and panoramic image sequences are taken simultaneously using a specially adapted stereo rig or by moving the tripod in carefully determined intervals to ensure consistent arc distance around a complete horizontal circle (Figure 3.2). The left camera position stays fixed about its nodal point while the right camera position traces the arc in successive steps at radius $r$ (where $r$ = the stereo base). The number of photographs taken is a function of the camera's field of view and the percentage overlap of successive shots.

**Figure 3.2. Overhead view of rotating panoramic and stereo-pair image acquisition**




In the testing of this prototype, two approaches were developed. The first approach consisted of a single tripod (without a nodal head) and camera setup with 1.5 metre stereo base separation. While this setup was simple in design, this approach was discarded since it introduced significant errors, as the tripod had to be continuously moved from the left and right camera positions by the operator to approximate stereo coverage. These errors posed heavy burdens on the software processing modules developed by the author, and reliable results could not always be obtained. However, this first attempt did provide invaluable test images that served as inputs in later refinement scenarios. The development of this first prototype was carried out in the summer of 2000. Early results from these first prototype tests also revealed the following:

- Normal case configuration of the cameras (two camera axes are parallel to each other and perpendicular to the base line) provided optimal object

coverage over convergent configurations (camera axes directed inwards towards the centre of the base line)

- Base line separation of 1.0 metres provided the optimal balance of ease of use and overlap coverage. In general, the longer the baseline, the more the geometry of the system improves for objects further away from the camera. Since object to camera distances in this research prototype application vary from 2 m to more than 300 m (in the typical urban setting), a stereo base line length of 1.0 m proved optimal in testing.
- A stereo base length greater than 1 metre proved cumbersome and unwieldy in real life testing scenarios and was deemed impractical by the author in an urban type setting.

Further, in the early prototype design stage of this research, stereo coverage was accomplished by manually moving a single camera tripod setup from the left to the right camera position. Not only did this introduce serious errors into the analysis due to the instability of the setup (as described above), it meant that the left and right views of a scene were not taken simultaneously. As a result of this time lag, the probability of the scene changing between shots was increased. This issue will be explored further in the stereo matching section of this chapter.

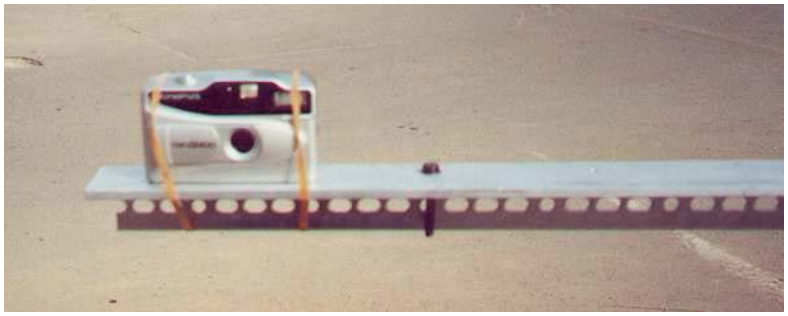
The second approach developed in the research consisted of a dual panoramic and stereo rig acquisition system. A working prototype is shown in Figures 3.2 – 3.5. The stereo bar consists of an aluminum bar (6 cm wide by 1.4 m in length) mounted on

top of a L-shaped steel support bar. The bar is then affixed to a modified tribrach to permit the desired 360-degree rotation. Counter weights were added to the short end of the rig for stability during rotation.

**Figure 3.3. Modified dual panoramic/stereo tripod**

**Figure 3.4. Left camera mount showing centre of rotation close to nodal point**

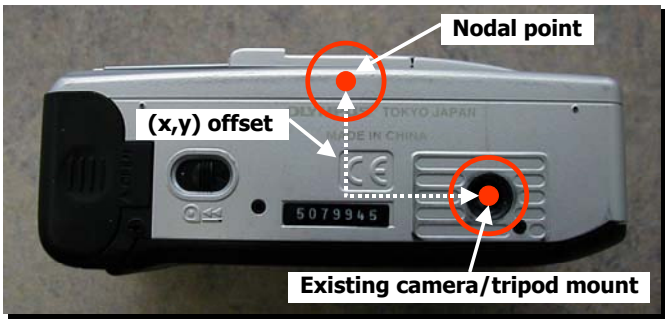
**Figure 3.5. Right camera mount**



Cameras were fixed to the stereo bar rig via the camera's internal tripod mounting mechanism and have a stereo base separation of precisely 1.00 metre. The stability of a camera mount is important since it can alter the relative orientation between the two cameras and distort subsequent distance calculations. The camera mounting holes were manufactured with mechanical tools of 25μm precision.

In total, 24 photos (ISO 100, standard DX-coded 35mm film) were acquired using the second prototype design at a testing location on an early morning of spring 2001 (Figure 3.7). Successive left stereo pair images have a consistent overlap (50%) to ensure effective mosaicking for subsequent panoramic warping. In this design, the right image pair is used exclusively for subsequent stereo model and space intersection calculations. Ideally, the left image should rotate about its nodal point (optical centre) to eliminate, through the use of a panoramic head, the potential for parallax in the sequence of left images. This necessitated the careful design of the stereo bar rig, since the mounted bracket for the left camera mount had to be slightly offset from the centre of rotation to account for the offset of the camera's tripod mounting mechanism (Figure 3.6).

**Figure 3.6. Nodal point offset for Olympus TripXB400 camera (bottom view)**



The introduction of parallax *within* the left stereo pair image sequence can make it difficult to stitch the sequence together (the stitching process is detailed below). Furthermore, the left and right optical axes should be parallel and perpendicular to the surface plane. In practice, a slight misalignment of either the nodal point or axes was unavoidable but can be tolerated. As such, a slight *x*-parallax was noticeable in the sequence of photos taken by the left camera for the test site. This did present some minor complications in panoramic image alignment.

**Figure 3.7. Left (left column) and right (right column) sequence of camera shots, of City Hall, Fredericton, NB, April, 2001**

Image 3

Image 3

Image 4

Image 4

Image 5

Image 5

Image 6

Image 6

Image 7

Image 7

Image 8

Image 8

Image 9

Image 9

Image 10

Image 10

Image 11

Image 11

### 3.3 Panoramic Warping

The 24 images scenes were processed, developed, and scanned commercially using the Kodak PhotoCD system to a digital image resolution of 1536 x 1024 (Kodak, 2001). The resolution was later re-scaled to 384 x 256 to reduce image file storage sizes (289 KB instead of 4611 KB each). This system preserves the full aspect ratio of the original image of 1½, and thus is suitable for photogrammetric measurement purposes. Scanning prints via a desktop flatbed scanner is not recommended due to the fact that 1) the prints are cropped at the photo developing source and thus the image coordinate system is difficult to establish for photogrammetric purposes; and 2) significant errors

can be introduced in the non-metric scanning process that are not addressed by the current calibration and processing software used for this research.

Each image was then converted from the proprietary Kodak format to the Tagged Image Format, or TIF. TIF is a public source file structure for digital images, and is considered a *de-facto* standard in image processing and is readable by the majority of image viewers available today. The TIF format was selected as the native format for image processing in this research due to 1) the wealth of information provided in the header file structure is advantageous; 2) 24-bit colour is fully supported (8-bits each for red, green, and blue), and; 3) the data structure is well known and documented. Disadvantages of the TIF structure include complexity of design for writing images, especially for simple read/write applications, and the large uncompressed file sizes. It should be pointed out that the latter can be overcome through the use of the proprietary LZW compression technology that is fully supported by the TIF image format.

Clearly, a pair of digital cameras could have been used instead of the traditional film based approach, and thus the time consuming process of scanning could have been avoided. However, high-resolution digital cameras come with a high-end price and thus deemed too expensive for the purposes of this research project. At any rate, the use of digital cameras instead of traditional film based cameras is fully supported in the hardware and software modules developed for this application.

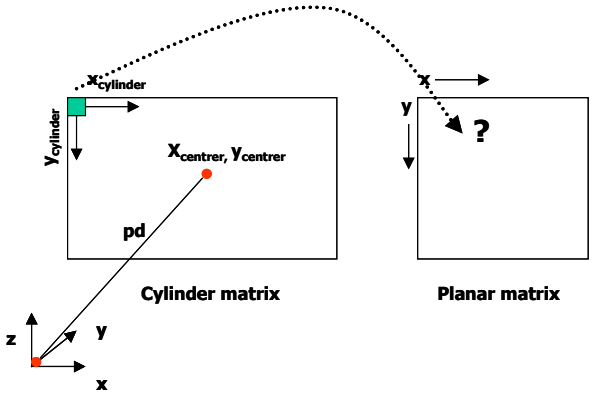
### 3.4 Projection of a plane to a cylinder

The left set of stereo pairs was then projected on a cylinder and stitched to form a complete mosaic using software developed by the author. The mapping of a plane to a cylinder is a well-known and understood geometric concept. In fact, there are numerous commercially available products capable of warping a sequence of overlapping images into a cylindrical panoramic image. However, for the purposes of this research, it was determined that these commercial products were not easily modified or adapted. Furthermore, the lack of specific details on precisely how the input images were being warped would present complications in subsequent stereo space intersection calculations.

The algorithm developed for plane-to-cylinder warping is as follows: given a pixel in the projected image, calculate the corresponding pixel location (and thus set of RGB brightness values) in the planar image, and copy the set of RGB values to the projected pixel location under consideration (Figure 3.8). If the pixel location calculated does not fall within the bounds of the original image, then the projected pixel value is assigned a zero value (RGB = [0,0,0]). The projected image is then cropped in the $x$ direction to remove any empty black space to facilitate the mosaicking process that follows. Each cell in the "empty" cylindrical projected matrix array must be visited exactly once in this way in order to obtain a fully representative output image. The following sets of equations are used to convert from planar to cylindrical $(x,y)$ coordinates:

$$
\begin{aligned}
x &= pd\,\frac{x'}{y'} + x_{centre} \\
y &= pd\,\frac{y'}{z'} + y_{centre} \\
&\text{where :} \\
x' &= \sin\left(\frac{x_{cylinder} - x_{centre}}{pd}\right) \\
y' &= \frac{y_{cylinder} - y_{centre}}{pd} \\
z' &= \cos\left(\frac{x_{cylinder} - x_{centre}}{pd}\right) \\
pd &= \text{estimated principal distance} \\
x_{centre}, y_{centre} &= \text{image centre (plane) in pixels} \\
x_{cylinder}, y_{cylinder} &= \text{cylinder coordinate s in pixels}
\end{aligned}
\tag{3.1}
$$

**Figure 3.8. Population of warped cylindrical matrix using planar matrix**




### 3.4.1 Resampling Techniques

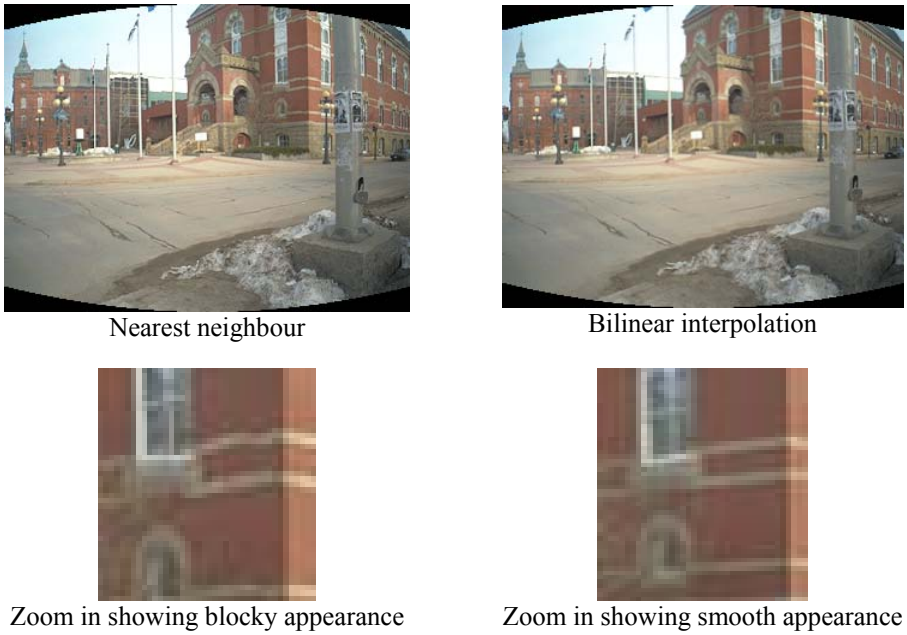
Unfortunately, the equation above does not yield exact pixel locations in the original planar image; that is, the *x,y* coordinates computed are not always integer numbers and thus do not fall exactly within the centre of the pixel. Therefore, some decision has to be made about what planar RGB values should be chosen for placement in the newly projected matrix array. Two resampling approaches were incorporated into

the warping process developed for this research: namely, nearest neighbour and bilinear interpolation. These contrasting approaches are described below.

#### 3.4.1.1 Nearest Neighbour Resampling

The nearest neighbour approach, one of the simplest resampling techniques commonly used in image processing, chooses the pixel that has its centre nearest the *x,y* point calculated in the planar image. This pixel, and its corresponding RGB values, is then copied to the projected pixel grid. Although this approach is computationally attractive and avoids having to alter the original input pixel values, nearest neighbour approaches tend to result in "blocky" and disjointed output images (Figure 3.9) since features may, in theory, be spatially offset up to ½ of a pixel. This is especially significant where warping is extreme.

**Figure 3.9. Nearest neighbour versus bilinear interpolation resampling techniques of warped test image (pd = 288 pixels or 27 mm), Fredericton, NB, April, 2001**



Nearest neighbour | Bilinear interpolation

Zoom in showing blocky appearance | Zoom in showing smooth appearance

### 3.4.1.2 Bilinear Interpolation Resampling

In contrast, bilinear interpolation takes three linear interpolations using the four pixels that share a side with the pixel corresponding to the calculated *x,y* location. This is a more sophisticated method since output pixels are assigned a set of synthetic RGB values. The process reduces computationally to a distance weighted average of the four neighbour pixels surrounding the pixel location of interest. This approach generates a smoother appearing output image and although it alters the original input values, this approach generated a superior output raster matrix. Based on these rudimentary practical tests, the bilinear approach was used as the principal resampling technique in this research.

From Equation 3.1, it is readily apparent that the estimated principal distance (*pd*) is the parameter that affects the magnitude of the cylindrical warping on the original image (Figure 3.10). In general, warping the input image sequence based on the "published" focal length or principal distance of the camera provided effective simulation of the panoramic effect. Calibration of the camera to precisely compute the principal distance of the camera was, in practice, not required for image warping.

**Figure 3.10. Image Distortion as a function of principal distance for image sequence 0, Fredericton, NB, April 2001 (Original image size 384 x 256 pixels)**

*pd* = 400 (358 x 256 pixels)

*pd* = 200 (306 x 256 pixels)

*pd* = 300 (347 x 256 pixels)

*pd* = 100 (218 x 256 pixels)

### 3.5 Image Alignment

Following the warping of each input image corresponding to the left camera position to simulate a panoramic image scene, each image within the sequence of left photo positions was aligned and blended in order to form a complete and seamless output mosaic.

Due to the amount of overlap between successive images in the panoramic sequence and the artificially introduced cylindrical warping, overlapping images manifested, in practice, only a simple *x*, *y* translation. That is to say that given, for example, image *a* in a sequence of *a*+*i* images, image *a*+*i* can be aligned with image *a* by translating it *x* pixel units in the negative *x* direction and *y* pixel units in the positive or negative *y* direction. In theory, since each interval of rotation was equal (or 360/12 = 30 degrees) during the image acquisition process, the *x* and *y* translation should have been consistent throughout each successive image pair. However, since the rotating stereo bar was not robotic or mechanized, the *x,y* translation was rarely consistent across the image sequence. Therefore, the translation across each overlapping image sequence pair had to be determined separately.

The determination of the translation parameters for each overlapping pair within the panoramic sequence of images reduces to an image-matching problem similar to that outlined in Chapter 2 of this thesis. In theory, the image sequence can be manually manipulated to form an output mosaicked image. However, this is a tedious process (11 image pairs!) and this author suggests that manual interaction detracts from the usability of the system. It was therefore decided that an automated approach was advantageous and therefore merited further examination. Two approaches were examined and implemented to generate translation estimates from the sequence of warped images: brute force correlation matching, and adaptive correlation matching. These are described in the proceeding sections of this chapter.

### 3.6 Panoramic Image Sequence Characteristics

There are three basic assumptions with respect to matching a pair of images within a sequence of warped panoramic images that assist in the design and development of an effective matching algorithm:

1. Image pairs manifest approximately 50% overlap and thus share similar features within each respective image. In this way, the prevailing assumption is that given a pair of images, there are corresponding features to match up in each.
2. The amount of offset or translation between image pairs is consistent across the entire overlap region.
3. Given a pair of matching pixels in corresponding images, the translation between these images in $x$ and $y$ can be readily derived.

While these assumptions hold for the purposes of estimating the translation between successive images within a sequence, it should be pointed out that these assumptions may or may not hold in more general image matching scenarios.

### 3.7 Brute Force Correlation Matching

The approach used here for brute force correlation matching incorporates image intensity information (a set of RGB values) to determine a set of matching locations in the overlapping areas of each image. Although the brute force technique did provide some reasonable results and was employed throughout the initial stages of the prototype design, the technique was deemed to be not sufficiently rigorous for the purposes of providing translation estimates and its full implementation was subsequently abandoned. However, the technique is presented here in order to provide context as to the difficulties encountered with aligning the panoramic image sequence automatically.

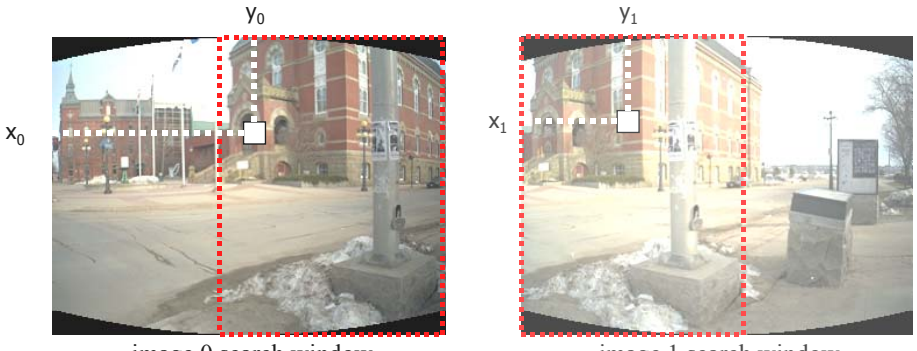
#### 3.7.1 Algorithm Description

Given a pair of images to be matched, the approach equates to a normalized cross correlation function in the form of the equation outlined in Equation 2.11. In the case of the notation from Equation 2.11., $x$ corresponds to image $a$ and $y$ refers to image $a+i$ in a panoramic sequence. This process is repeated until all successive image pairs within the panoramic image sequence have been examined. Interestingly, although a match should, in theory, be obtained when the cross-correlation value reaches 1, even user determined

matches can deviate from this ideal. In the initial testing of this prototype, cross-correlation scores ranged from 0.65 to 1.0 for image sequence 0 and 1 for 10 user defined test matches. This is largely due to differences in the corresponding RGB values for each overlapping camera shot as a result of sun illumination, exposure, and commercial developing differences.

In order to compute the cross-correlation function, a template window is shifted pixel by pixel across a larger search window, and in each position the cross-correlation coefficient between the source template window and the corresponding region of the search window is computed (Figure 3.11.). The area of interest within the left image is constrained through the assumption that its right matching image overlaps approximately ½ its width. Therefore, given a left image of *w* by *h* pixels, at most *w/2* x *h* pixels (corresponding to its right most half) must be examined. The same assumption holds for the next image in the sequence, except its corresponding left half is examined.

**Figure 3.11. Constrained search windows for image sequence 0 and 1 showing kernels centred at ($x_0$,$y_0$) and ($x_1$,$y_1$) respectively, Fredericton, NB, April 2001.**



The maximum of the cross-correlation defines the position of best match between the source template and the search template window. This is an exhaustive technique since for a complete solution, all pixels in each search region must be examined (although as noted below, thresholding can reduce the number of pixels examined). Once a match has been determined, the translation in $x$ and $y$ can, of course, readily be calculated.

### 3.7.2 Brute Force Approach Refinements

Key considerations for effective matching are: 1) selecting an appropriate size of the neighbourhood surrounding the pixel of interest; 2) incorporating pre-estimates of the translation in $x$ and $y$, and; 3) using a correlation threshold to denote a probable match and thus halt the matching process pre-maturely. The size of the kernel surrounding the pixel of interest must be chosen carefully: it must be small enough to uniquely characterize the region of interest, and it must be large enough to reduce the probability of image noise skewing the analysis. Unfortunately, it is non-trivial to decide upon an appropriate kernel size, as noted in Chapter 2. Factors influencing this decision include the resolution of the image, the quality of the image (level of noise), and the specific feature under consideration. In practice, choosing an appropriate kernel size is largely a case specific exercise requiring, to a large extent, experimentation through trial and error. For the purposes of this research, a 7 by 7 pixel template provided the best translation estimates using the brute force technique.

It is not unexpected that the quality of a match can be improved through the incorporation of translation pre-estimates into the analysis. This has the effect of narrowing the number of pixels to be considered (thus costly floating point correlation calculations) through the assumption that the match is most "likely" to occur in close proximity (for example within +/- 20 pixels of a given pre-estimate in $x$ and $y$) to a user supplied pre-estimate. This is especially significant in the $y$ direction, as, in practice, each image was rarely offset from its overlapping neighbour by more than +/- 10 scanlines. As discussed in Chapter 2, this reduced the 2-dimensional search problem to a more manageable 1-dimensional search problem without re-projecting the images to the common epipolar projection. A similar approach also applies to the $x$ direction; however, as previously noted, offsets in the $x$ direction were not consistent across successive overlapping images. As a result, supplying pre-estimates in the $x$ direction are not as effective for reducing the total search time.

The use of a correlation threshold can also serve to limit the extent to which the images are searched. A correlation threshold works as follows: given a computed cross-correlation measure between two images, a match (and thus set of translation parameters) is obtained when the measure is greater than or equal to a pre-established correlation threshold. In this way, the remaining pixels are not examined and the process halts. However, thresholding does not *guarantee* a match is found in shorter time period or even that the match reflects the actual translation between the images. The threshold is chosen to maximize the likelihood that a given cross-correlation value actually represents a true match. Unfortunately, similar to the derivation of an appropriate kernel size, there

is no universally accepted method for determining an appropriate threshold value. For the purposes of this exercise, a threshold value of 0.85 provided the best matches possible using the brute force correlation technique.

This initial brute force correlation strategy, despite experimentation with various kernel sizes, pre-estimates, and threshold values, did not provide reliable translation estimates, and in the case of 8 of the 12 image pairs, failed to find a match at all. This author suggests that this was due to 1) difficulty in estimating an appropriate threshold correlation value and thus too many or too few good matches were found; and 2) the quality of a good match does not necessarily correspond to a high cross-correlation score. Condition 2 is largely a result of homogeneous pixel regions that manifest high correlation scores throughout a broad area. Examples of homogeneous pixel regions include grass, roadways, and building facades. In this way, the target pixel neighbourhood is simply not distinct enough to uniquely characterize that particular location, and thus the algorithm returns with poor matching estimates, even though the correlation score approaches optimality (Figure 3.12.).

**Figure 3.12. Homogeneity of pixel regions in image matching**

Image 0

Image 1

Zoom in on street

Zoom in on street

It is worth noting the choice of similarity measure (cross-correlation, colour separation) did not improve or worsen the matching results, either in terms of processing time or quality of match. A full description of the results obtained from the various matching approaches, including the adaptive matching approach outlined below, is documented in Chapter 4 of this thesis.

### 3.8 Adaptive Matching Approach

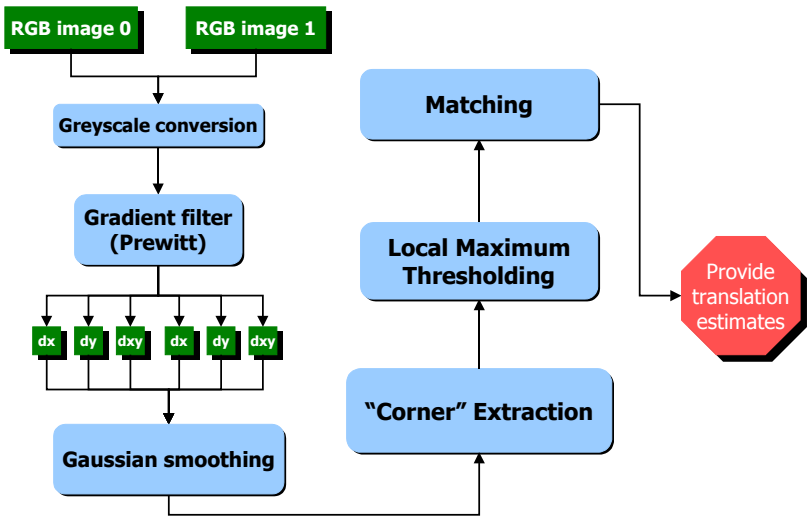
The adaptive matching approach, in contrast to the brute force matching technique, relies on the extraction of a subset of features within each of the images that are then assessed to determine the most appropriate match. This has the effect of reducing the total amount of comparisons required while at the same time providing a good set of initial pixel locations to begin the search for matches. In its most basic form, the adaptive matching approach attempts to extract a subset of pixel locations from each image using a neighbourhood pixel gradient technique to extract “corners” and then compares these

extracted pixel locations for the most likely matches using a combination threshold and classical correlation based matching approach.

Since each "candidate" match denotes a translation estimate, matches with translation estimates within a pre-defined neighbourhood (say, for example, 3-5 pixels in x and y) are placed in the same set. The best match and thus the best translation estimate is selected as the set with the most number of matches found. The prevailing implication, of course, is that the number of matches found for a given translation neighbourhood provides direct evidence for or against the "goodness" of a match.

The assumption here is that by extracting a subset of high contrast and distinctive pixel locations, the problems associated with homogenous pixel regions can be overcome by attempting to avoid them altogether. The process developed in the research for deciding upon and then matching a subset of pixels in shown in Figure 3.13. The extraction of high contrast pixel locations is computed using a Plessey type operator specifically modified for this research and originally developed by Harris (1987) and later improved by Harris and Stephens (1988). These approaches use edge information extracted from an input image to denote possible corner pixels. Corners can be characterized not only in the sense of intersections of image lines, they capture corner structures in patterns of intensities. Such features have been shown to be stable across image sequences and are therefore useful as aids to track objects (Noble, 1988). Therefore, the prevailing assumption here is that corner pixels are useful as initial matching points. A detailed description of the adaptive matching algorithm components is presented in the following sections.

**Figure 3.13. Overview of adaptive matching algorithm components for image stitching**




### 3.8.1 Algorithm Description

The image pair is first converted to greyscale to simplify the proceeding steps in the algorithm by reducing the total number of two dimensional arrays requiring processing (1 array per image instead of 3 arrays per image). The conversion from RGB true colour to greyscale is performed as follows:

$$GSP = (0.20 \times R) + (0.60 \times G) + (0.20 \times B) \qquad (3.3)$$

where $GSP$, $R$, $G$, and $B$ values correspond to the greyscale, red, green, and blue component values respectively. It is worthy to note that the conversion to greyscale does

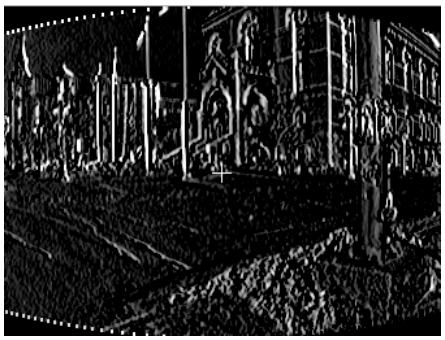
not impact the quality of the output panoramic image; rather, the conversion to greyscale is a processing step that is hidden to the user.

Next, *x* and *y* directional gradients are applied to the greyscale image pair through the use of the follow set of templates:

$$dx = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad dy = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad dxy = \sqrt{dx^2 + dy^2} \qquad (3.4)$$

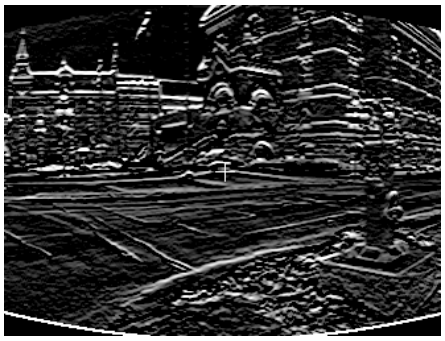
These templates are simple yet effective techniques for extracting both the horizontal and vertical edges within an image and are used in conventional edge extraction operations. In fact, the gradient *dxy* is commonly referred to as the Prewitt edge detection approach (Prewitt, 1970). Edge information within an image corresponds to any significant change in digital number value across the image. The use of the templates above amplifies any edge information found within the images (Figure 3.14).

**Figure 3.14. Vertical, horizontal, and magnitude outputs for image 0 and 1, City Hall, Fredericton, NB.**

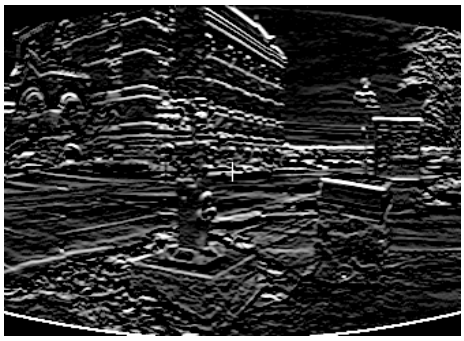


vertical edges extracted

vertical edges extracted

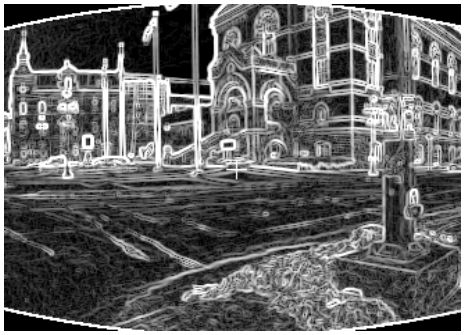


horizontal edges extracted



horizontal edges extracted

magnitude of the gradient (dxy)



magnitude of the gradient (dxy)

Clearly, in the absence of image noise, the edges extracted above correspond directly to a feature found within the image. Unfortunately, the edges extracted above can contain significant noise that can lead to erroneous corner information being extracted in subsequent steps. Therefore, following the extraction of horizontal and vertical edges for each image, each output matrix is smoothed using a 2-dimensional Gaussian function in order to reduce the effect of image noise in the analysis. The Gaussian function reduces to a weighted average filter technique whereby pixels further away from the central pixel under consideration are given lower and lower weight (and vice-versa). A 2-dimensional Gaussian function can be expressed as follows:

$$G(x,y) = \sigma^2 e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \qquad (3.5)$$

where $\sigma = 1$. The Gaussian function is a well known operator in image processing for reducing high frequency image information. This function can be reduced to a more computationally efficient solution by separating the Gaussian into two 1-dimensional convolutions and then taking the magnitude of the result ($\sqrt{G(x)^2 + G(y)^2}$) for each pixel. As suggested by Parker (1996), this provides for as effective an approximation as the more costly 2-dimensional Gaussian approach. For the purposes of this research, a 2-dimensional Gaussian was applied using a 5 x 5 kernel window which, in initial testing, provided the optimal balance between noise suppression and excessive smoothing.

It is important to note that there are many other edge detection algorithms presented in the literature, such as the Canny (1983), Beaudet (1987) and Kitchen and Rosenfeld (1982). In fact, many may provide superior results with respect to edge detection over the rudimentary technique employed in this research due to their improved error rate (respond to all and only edges), localization (actual edge and detected edge offset minimal), and response (identify single and multiple edges appropriately). However, it is important to point out that the goal of this exercise is to provide a good set of high contrast pixel locations and not necessarily to extract all possible locations (as is largely the case with conventional edge extraction techniques). As such, the approach used in this research performs effectively based on the results of the stitching process.

While edge information alone does provide some benefits with respect to providing a set of initial high contrast pixel regions, the edge detection approach presented thus far fails to reliably detect the corners and intersections within an image that tend to have the highest information content (Noble, 1988). Unfortunately, even the

more elaborate and sophisticated techniques fail in this regard. Further, the continuous nature of edge pixels introduces homogeneity along the edge direction and thus makes them unsuitable for use as matching points.

There are numerous approaches for extracting corner information within a digital image, including Kitchen and Rosenfeld (1982), and Noble (1988). A Plessey (Harris, 1987) type approach was selected for incorporation into this research due to the relative low computational cost of implementing the algorithm as well as strong citation record of the approach in the computer vision and image processing literature (Harris, 1988; Noble, 1988; Zhang *et al.,* 1995). Harris and Stephens (1988) considered a slightly modified version of the original Plessey corner detector. From Moravec's (Moravec, 1977) work as well as Barard and Thompson's (Barnard and Thompson, 1980) investigations, they defined a measure for extracting corners based on the following operator:

$$C = \begin{bmatrix} dx^2 & dxdy \\ dxdy & dy^2 \end{bmatrix}$$
$$R = |C| - k(trace^2 C), \text{ where } k = 0.04 \qquad (3.6)$$
$$R = dx^2 dy^2 - (dxdy)^2 - k\left(dx^2 dy^2\right)^2$$

where *dx, dy,* and *dxy* denote the smoothed (2D Gaussian with sigma = 1) "edge" extracted matrices approximated using a discrete 3 x 3 Prewitt template as above, and *R* refers to the output "cornerness" matrix. Positive values of *R* denote corners, whereas negative and near zero responses denote edges and homogeneous regions respectively. The value *k* represents a threshold that is useful for providing discrimination against high

contrast pixel step edges. While Harris and Stephens (1988) did not explicitly denote acceptable values for $k$, subsequent studies have revealed that when $k = 0.04$, the optimal number and quality of corners can be detected (Zhang *et al*., 1995).

In its most basic form, $C$ characterizes the structure of the intensity values making up the image. Since $C$ is in fact symmetric, it can be transformed using a principal component transformation. The derivation of the eigenvalues are given by the solution to the characteristic equation:

$$\begin{vmatrix} C - \lambda I \end{vmatrix} = 0$$
$$\begin{vmatrix} dx^2 - \lambda & dxdy \\ dxdy & dy^2 - \lambda \end{vmatrix} = 0 \qquad (3.7)$$

with $I$ being the identity matrix. Therefore, $C$ can be reduced to the following:

$$C = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad (3.8)$$

where $\lambda_1$ and $\lambda_2$ refer to the eigenvalues of the matrix $\boldsymbol{C}$, with $\lambda_1 > \lambda_2$. In this way, this author suggests although implicit in their derivation of a corner detector, Harris and Stephens (1988) avoided the explicit eigenvalue decomposition of $\boldsymbol{C}$. Accordingly, there are three possible cases with respect to the computation of $\boldsymbol{C}$ and $R$ above:

1. Both *dx* and *dy* are small; thus, the region is considered homogeneous (local auto-correlation function is flat) with $\lambda_1 = \lambda_2 = 0$ and thus $R = 0$
2. Either *dx* or *dy* is large; thus, the region is considered "edged" (local auto-correlation function is ridged shaped) with $\lambda_2 = 0$, $\lambda_1 > 0$ and $R$ is negative
3. Both *dx* and *dy* are large; thus, the region is considered a "corner" (local auto-correlation function is highly peaked) with $\lambda_1 > \lambda_2 > 0$ and $R$ is positive

While *R* could, in theory, be thresholded to produce an isolated output corner matrix, in practice, this does not provide for effective results. In fact, the *k* parameter does provide for a type of threshold for discerning corner information. In a similar methodology employed by Canny (1983) and implemented by Parker (1997), the output *R* raster matrix must undergo non-maximum suppression in order to remove neighbourhood pixels (3 x 3 neighbourhood) that are not local maximums. This is similar to an adaptive threshold but has the added benefit of thresholding the image matrix *R* based partly on the direction of the image gradient.

Non-maximal suppression implies that the pixel under consideration must have a larger gradient magnitude than its neighbours in the gradient direction (Parker, 1997). After computing the horizontal and vertical gradients of *R* using the Prewitt approach developed earlier in this thesis, the non-maximum suppression algorithm works as follows (assuming a moving 3 x 3 neighbourhood kernel across the entire *R* matrix):

1. Starting at the central pixel, move in the direction of the gradient until a new pixel A is found
2. At the central pixel, move in the direction opposite of the gradient until a new pixel B is found
3. If in moving from pixel location A to B the gradient value at the central pixel is higher than the gradient values at both A and B, then denote a corner pixel

In cases where the gradient direction at the central pixel is not perfectly horizontal or vertical, a linear interpolation is applied to estimate an appropriate gradient value at the prescribed location. The results of the corner extraction process are shown below (Figure 3.15.). It is worthy to note that even in relatively texture-less and featureless regions, corner information can often be extracted.

**Figure 3.15. Extracted corners using the adaptive corner extraction technique for test image sequence 1, City Hall, Fredericton, NB.**

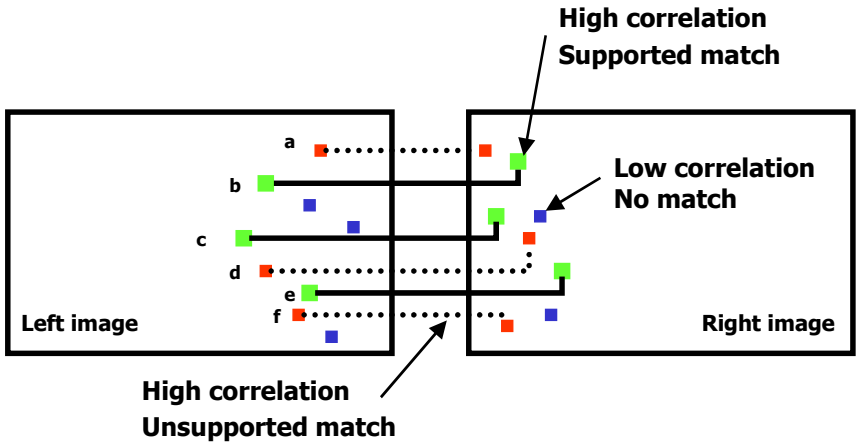
extracted corners denoted as squares

Once corners are extracted for each image pair (corresponding to a set of pixel coordinates), each corner location found in one image is compared to the overlapping image's corner locations using a cross-correlation based similarity measure applied to the original image intensity values. A relatively low correlation threshold is applied (0.60) such that the current match under consideration is classed as a "candidate" match if it is above this threshold. These candidate matches are then further classified into sets using the derived translation estimate for each match (Figure 3.16). Thus, candidate matches are only placed in the same set if they predict similar translation.

User defined pre-estimates are also applied so as to restrict the total number of comparisons required. Further, the area to be searched can be restricted as outlined above in Figure 3.11. This entire process continues until all corner combinations have been examined.

Finally, the best match and thus translation estimate between overlapping images is chosen as the group with the most members and consequently the most support for the translation. Processing continues until all image pairs have been examined, including the last and the first images within the sequence. If all sets corresponding to translation estimates are null, such that no possible candidate matches under consideration are greater than the *a priori* defined threshold, the process is re-started with a lower threshold (new threshold = old threshold – 0.05). While lowering the threshold may seem entirely arbitrary, it does provide for effective results and, in the cases of difficult to match image pairs, provides at the very least an effective estimate.

**Figure 3.16. Candidate matching support concept**



The above figure exists where *{b, c, e}* form the set with the most support for the estimated translation, and *{a}*, *{d}*, and *{f}* form sets with lesser support for their respective predicted translation. There are several key advantages of the adaptive matching approach over the brute force correlation approach outlined earlier in this thesis:

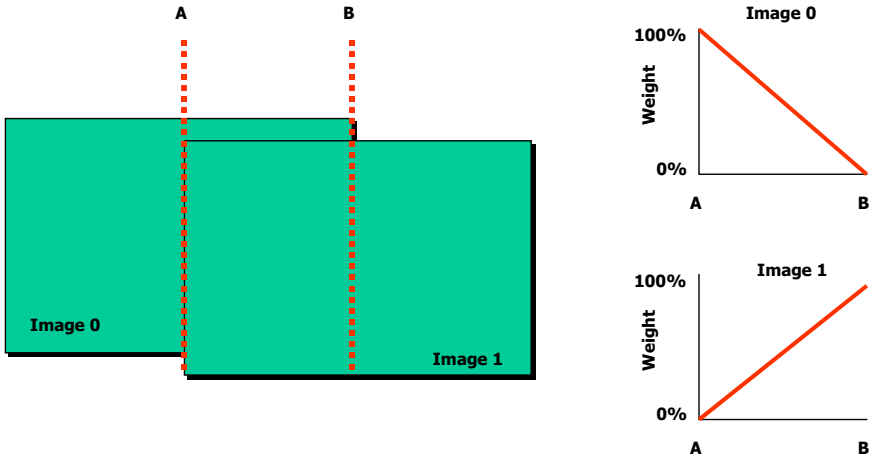
- There appears to be less significance associated with the selection of a correlation kernel size; a 7 x 7 kernel provided similar results to that of a 5 x 5 and 9 x 9 kernel size
- Due to the small number of comparisons required, the threshold can be adaptive in nature; that is, if evidence is not conclusive for a given set of translation estimates, then the analysis can be repeated with a lower correlation threshold until sufficient evidence is generated for a specific translation estimate.
- Computation times appear to be significantly shorter for a given image pair
- The translation estimates computed are better in quality and the process is more rigorous (ie. it works effectively for more image pairs in testing)

### 3.9 Blending and End-to-End Alignment

Following the extraction of the appropriate translation parameters for each image pair, the images are automatically mosaicked into a seamless output image using a linear

blending function. This function feathers the images based on the translation offsets estimated by the previous alignment process. Based on the example provided below (Figure 3.17.), the calculation of the appropriate RGB intensity values using this technique for the overlap region is as follows: along the **A** axis, the output blended mosaic is assigned 100% of image 0's intensity values, and 0% of image 1's intensity values. Similarly, along the **B** axis, the output blended mosaic is assigned 0% of image 0's intensity values and 100% of image 1's intensity values. The pixel location in between are assigned values based on a linear gradient between the **A** and **B** axes.

**Figure 3.17. Linear function for blending offset image pairs**




Not unexpectedly, this process continues until all images have been included in the output mosaic. Unfortunately, due to the accumulation of 1) small errors in the *y*-translation estimate, and; 2) camera movement during acquisition across the panoramic image sequence, the left edge of the first image making up the mosaic is unlikely to match up perfectly with the extreme right edge of the last image in the sequence. In order

to counteract this vertical drift, the image mosaic is warped using a 1st order polynomial equation in the form:

$$x' = a_0 + a_1 x + a_2 y + a_3 xy$$
$$y' = b_0 + b_1 x + b_2 y + b_3 xy \qquad (3.8)$$

where *x'* and *y'* refer to the new output mosaic, *x* and *y* denote the original mosaic, and $a_0 \ldots a_3$ and $b_0 \ldots b_3$ represent the model coefficients. The model coefficients can be derived easily due to the fact that the accumulated *x* and *y* translations have been previously derived through the image matching/alignment process and are thus assumed as known quantities. In this approach developed by the author, image 0 within the sequence is blended into the mosaic at both the left and right sides of the mosaic. Then, using the polynomial warping function outlined above and a bilinear resampling technique, the vertical drift is removed. Finally, the ends are clipped and blended to form a complete and seamless 360 degree panoramic image (Figure 3.18).

**Figure 3.18. Seamless output mosaic as the product of automatic warping, alignment, blend, and end-to-end alignment for City Hall, Fredericton, NB, test location**

### 3.10 Visualization

Numerous panoramic viewers are currently available that can display a panoramic image so as provide the effect of complete 360 degree viewing. However, as is the case for panoramic warping, alignment, and blending, only minor modifications are possible with a commercial viewer and its corresponding Application Programming Interface (API). Further, the core implementation details are hidden to the user and the majority functionality offered by these commercial environments are largely non-computational in nature. This presents obstacles for effective integration with not only space positioning, but with design modifications in the future.

From a computational perspective, a panoramic viewer must be able to efficiently re-project the simulated panoramic image onto a planar surface so as to offer almost unlimited viewing perspectives in a complete 360 degree horizontal rotation. Mathematically, this reduces to the inverse problem documented in panoramic warping section of this thesis.

The visualization interface designed by the author, while continually a work in progress, allows the user to scroll horizontally and vertically around a given viewpoint (Figure 3.19). It is important to point out that the goal of this exercise is to provide proof of concept for the design, development, and implementation of a spatially enabled panoramic viewing environment, and not exclusively a “commercially” enabled software product.

**Figure 3.19. User interface for visualization of panoramic image**

Map view

Panorama Visualization

## 3.11 Space Positioning

Although the visualization of the panoramic image provides for an effective virtual reality simulation, the determination of object distances from the camera within the panoramic image is the focus of this section and the main thrust of this research. The incorporation of the spatial component to the panoramic viewing environment is achieved through the use of the as yet unused right image pair of each left image making up the panoramic sequence. This essentially reduces this stage of the research to that of a non-metric close-range photogrammetry problem. A further complication is introduced due to the fact that the prototype developed should allow for simple user defined "point-and-click" queries within the panoramic environment. This would provide the user with the ability to automatically determine the object-to-camera distances of any feature in a given panorama. This valuable distance information can then, in theory, be related to a real

world coordinate through the use of a supplemental positioning device, such as a commercially available handheld GPS unit, and find use within a GIS.

As outlined in Chapter 2, at its most basic level, object distances can be calculated through the knowledge of the image position of the object in the left and right images, the principal distance of the camera, and the stereo baseline separation between the left and right cameras (2.1, 2.2). This condition, of course, can only hold when the optical axes of both cameras are parallel to each other and exactly perpendicular to the stereo baseline axis. This photogrammetric "normal case" is computationally advantageous and extremely simple to implement but can suffer from poor accuracy and reliability, especially in the case of non-metric inputs. However, irrespective of how object distances are calculated, given the assumption that the normal case could provide sufficient accuracy for the purposes of this specific application, two key problems must be solved:

1. Panoramic coordinates must be related back to the original left image coordinate values
2. The corresponding location of the feature of interest must be found in the right image through an automated process

In the more general photogrammetric case, the problem of determining object distances is further complicated since both the interior and exterior orientations for each camera position must be determined. Camera calibration yields interior orientation parameters and corresponding exterior orientation parameters that can, in theory, produce significantly higher levels of accuracy in data reduction computations compared with the

normal case presented above. This is especially significant given the non-calibrated camera and tripod setup used in the photo acquisition process of this research.

While both approaches for determining the spatial position of an object within an image are computationally quite different, both techniques must solve the fundamental problems of relating the panoramic image coordinates back to the original left image as well as the automatic determination of the corresponding location of the same feature in the right image. Not surprisingly, the latter shares significant similarity with the image alignment/matching process previously documented in this chapter.

### 3.11.1 Conversion to Original Image Coordinates

The conversion from a user selected panoramic image coordinate back to the original image coordinate is, perhaps not unexpectedly, trivial in concept and can be achieved by reversing the panoramic imaging process developed above. This highlights the significance of developing, rather than using, a set of panoramic processing software modules. Quite simply, since the implementation details are hidden in the commercially available software packages, it is virtually impossible to relate back to the original imagery. In summary, the conversion to original image coordinates involves the following:

- Applying a reverse polynomial warp
- Determining the appropriate cylindrically warped image containing the coordinate
- Applying a reverse cylindrical warp to revert back to the original image

Further, an offset in the *x* and *y* direction is applied to account for any cropping of the images that was performed in the original panorama creation process to remove unwanted black null areas.

### 3.11.2 Stereo Matching

Once the appropriate pixel location has been computed for the left image of the stereo pair, the corresponding right location is found using an automated approach similar to that of image alignment and stitching presented above. As outlined in Chapter 2, the correspondence problem or image matching problem is non-trivial to solve and unfortunately, there exists no single solution giving optimal results under all possible circumstances.

Thus, while many techniques are currently presented in the literature, the approach used here consists of a combination of the conventional cross-correlation approach and the adaptive "corner" extraction approach documented earlier in this thesis. A combination approach was developed since the technique used previously for panoramic stitching did not provide reliable matches in initial testing. This was likely a result of the fact that this approach relies on a high contrast and well-defined initial pixel region to increase the likelihood of a strong match in the corresponding overlapping image. However in this scenario, the user, and not a computer-based algorithm, selects the initial object in the image with which to match. As such, the feature may or may not be a high contrast location. In fact, this author suggests that users are *more* likely to select

features that are homogeneous, such as building facades and roadways. While this concept may appear be to somewhat puzzling, this author suggests that it is perhaps not unexpected since users identify with macro-scale shapes in scenes (roadways, buildings), rather than specific micro-scale object features (such as corners and/or edges).

#### 3.11.2.1 Stereo Pair Characteristics

Similar to the process of image alignment, there are 3 key assumptions that can shed some light into the process of extracting reliable matches:

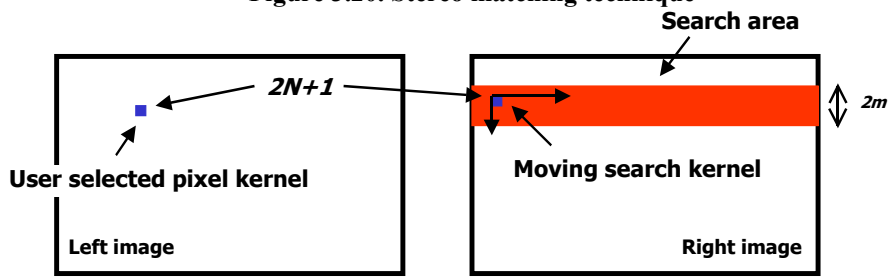
1. Enforcement of the epipolar constraint is not necessary in this application since, in practice, each stereo pair is no more than 3-4 scan lines offset in the *y* direction, and thus the matching can be reduced to a 1-dimensional search problem from the outset. This is advantageous since matching is performed on-line and computational costs can be high if not adequately constrained.
2. Due to the variability of object-camera distances, overlap can vary between the image pair from 0-100% for objects close and far away from the camera respectively. This can present limitations in that a match may not be possible for objects in close proximity to the camera. Similarly, the stereo separation between objects far away from the camera approach the measuring accuracy of the system.
3. Although the use of candidate matches and evidence of support provide very effective results in image alignment, this same technique can not be applied in the

case of stereo matching since only a single pixel location is of interest in the left image.

### 3.11.2.2 Algorithm Description

The basic algorithm for the matching of a user supplied pixel location in the left image with the right image is as follows: using a kernel of size *2N+1* and starting *m* units in the negative *y* direction (upwards) and at the most extreme lower *x* value, move the kernel across the image and compute the normalized cross-correlation between it and the same sized kernel surrounding the user defined pixel location in the left image. The process is terminated when the kernel either reaches the end of the image array or if it surpasses *m* units in the positive *y* direction (Figure 3.20). The arbitrary parameter *m* is selected *a priori* in order to minimize the amount of comparisons required. The position of maximum correlation defines the best estimate of the match. As noted earlier, a high correlation value may not necessarily denote the best possible match.

**Figure 3.20. Stereo matching technique**



#### 3.11.2.3 Algorithm Refinement

In order to reduce the potential of an erroneous match, two additional strategies are employed. First, a correlation threshold is applied to the analysis such that if the obtained maximum correlation value is not over a given lower threshold, then the *2N+1* kernel size is expanded one unit (*N=N+1*) and the process re-starts. The prevailing assumption here is that the original kernel size was not large enough to contain distinctive pixels and thus by enlarging its size, potentially more distinctive pixels can be included in the analysis. In practice, while this additional strategy can provide some useful matches, unfortunately, this approach fails in broad homogenous pixel regions. If in re-starting the analysis with a larger kernel size no further matches are detected, then the process halts and the second strategy is employed to find a possible match.

The second strategy employed involves the use of the adaptive corner extraction procedure developed earlier in this thesis. Specifically, this approach is used when all previous attempts at matching fail. Using the same approach for extracting "corners" or high contrast pixel locations developed above, corner information is extracted for each image. Next, the closest corner to the original user selected pixel is determined using an Euclidean function. This closest corner pixel then becomes the pixel of interest for determining a distance estimate. As such, this new corner location is compared with corners in the right image using a normalized cross-correlation technique. The corner combination with the highest correlation score is selected as the most probable match and thus the left and right image coordinate locations are considered found. Similar to the stitching process documented above, it is beneficial to apply a lower correlation score

threshold to reduce the likelihood of the algorithm selecting an erroneous "optimal" match. In this application, a lower threshold of 0.60 and a kernel size of 7 x 7 pixels ($N$=3) provided adequate results.

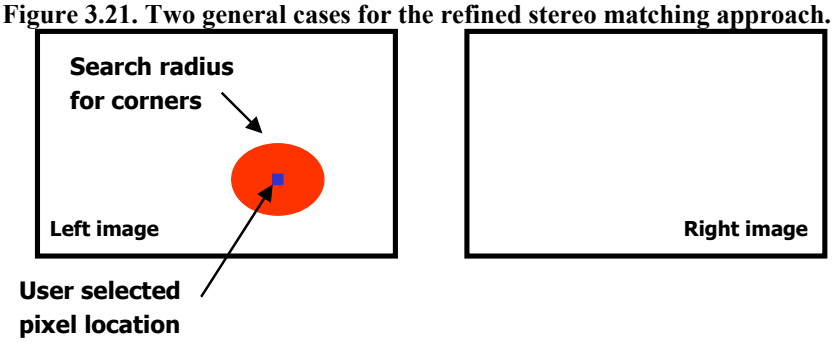
In this way, the prevailing assumption is that the closest corner pixel is in close enough proximity to the user selected pixel location and thus from a spatial location point of view this corner closely approximates the difficult to match pixel location. Clearly, a threshold is useful here to limit the proximity distance so as to maximize the potential that the extracted corner information actually represents the original user selected location. If no corner information can be found within this threshold proximity distance, then, for the purposes of this application, no match can be found and the process fails. In this analysis, a distance threshold of 20 pixels (5% of image height) provided the satisfactory results.

In the case where no corner match can be found above the *a priori* defined correlation threshold, then the algorithm selects the next closest corner to the original user selected pixel in the left image and the process repeats itself. If such a corner does not exist, or if no match can be found within the *a priori* defined correlation threshold, the stereo matching process halts without a match.

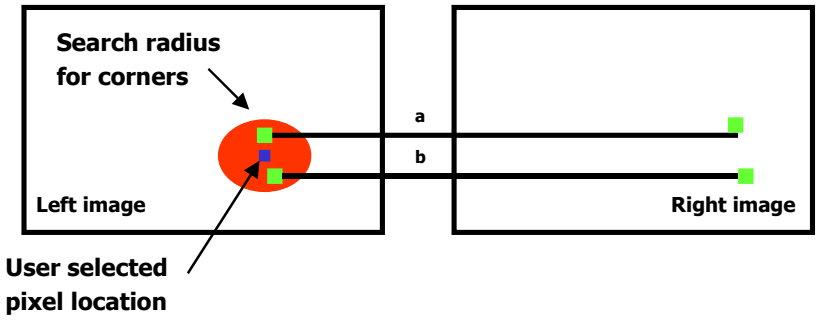
Figure 3.21 illustrates the two general cases commonly found in the refined stereo matching approach. In 3.21a, no corners are found surrounding the user defined pixel of interest, and thus the process fails since no match can be found. In 3.21b, at least one corner is extracted in close proximity to the user defined pixel location. The closest corner is then compared with corners extracted in the right image to determine the best

match. If no appropriate match can be found, then the next closest corner is processed in the left image, and so on.

**Figure 3.21. Two general cases for the refined stereo matching approach.**




**a. No corner information found**




**b. Two corners found and compared**

### 3.11.3 Calculation of object depth

Following the determination of the image coordinates of the object of interest in the left and right image, the parallax can be readily calculated and introduced into equation 2.1 to calculate the distance of the object from the camera system baseline. The "published" focal length and camera stereo baseline separation are used as known entities in the analysis, and thus no object space control is necessary.

This approach, of course, does not take into account the interior or exterior orientations of the cameras and stereo system setup respectively. As mentioned previously in this thesis, the normal case does present limitations since it does not take into account the inadequacies of the cameras and camera setup. This is especially significant in the case of non-metric inputs. The calibration of the camera and camera setup used in this thesis forms the focus of the next section.

### 3.11.4 Camera calibration

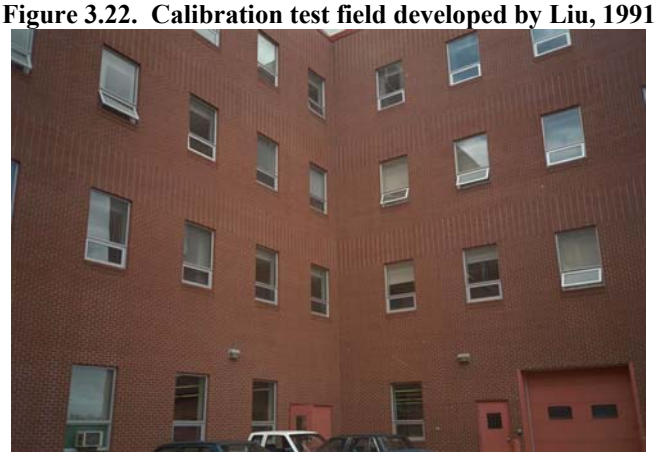
The necessity of camera calibration in non-metric high precision close-range photogrammetry is well established. Camera calibration typically involves the determination of the interior and exterior orientations of the camera and stereo-rig setup respectively. Interior orientation establishes the geometrical relationship between the perspective centre and image plane (principal point, principal distance), while the exterior orientation defines the position and orientation of the image in object space (*X, Y, Z* object space coordinates, and *x*-tilt, *y*-tilt, and swing) (Derenyi, 1996). The process of relative orientation is one aspect of exterior orientation that establishes the orientation of one camera to the other, and thus is based in a local coordinate system (in this case with origin at the left stereo camera position).

The approach used to calibrate the camera and stereo system setup is as follows: 1) a new set of photographs were acquired at a testing range designed explicitly for camera calibration so that the interior orientation of the camera could be determined; 2) these interior orientation parameters were used as known entities to facilitate the

determination of the relative orientation parameters for one of the previously collected stereo pairs. Thus, a pre-calibration scenario was used in the development of this prototype; that is, the camera was calibrated prior to use to determine the appropriate camera calibration parameters.

Five photographs were acquired of a geodetically surveyed test range using the same camera and commercial development process used for acquisition stage above (Figure 3.22). The test range consists of 61 3-dimensional points on 2 wall planes that have been previously surveyed to acceptable levels of accuracy (95% confidence level) for the purposes of camera calibration (Liu, 1991). Using an iterative Direct Linear Transformation (DLT) calibration methodology (Heikkila and Silven, 1998), the principal distance, $x$ and $y$ principal point offsets ($x_o$, $y_o$), as well as radial and tangential distortion coefficients ($K_1$, $K_2$, $P_1$, $P_2$) were calculated for each photograph. This modified DLT technique has the added benefit of directly modeling lens distortion parameters in the overall interior orientation solution and is suggested to provide high geometrical accuracy (Heikkila and Silven, 1998).

**Figure 3.22. Calibration test field developed by Liu, 1991**



The mean values of the principal distance, principal point offsets, and distortion parameters of the five calibration images for the left and right cameras were used to determine the relative orientation of each stereo pair making up the panoramic sequence from a selected test location (City Hall, Fredericton, New Brunswick) through the direct analytical calibration approach (dependent pair relative orientation) using 16 manually selected overlapping image points. The computed interior and exterior orientation parameters were then treated as known quantities in a set of collinearity equations to determine the distance of the object from the camera station through space intersection computations. Distance calculation results using a calibrated and non-calibrated setup are presented in the following chapter of this thesis.

### 3.11.5 Determination of Bearing

Along with distance, bearing information is critical for the effective integration in a map-based or GIS environment. In the implementation of this prototype, the x-intercept in panorama pixel units corresponding to the Northern direction is determined first. Since the total panorama image size is known, a conversion can be applied to convert from normal panoramic image coordinates to bearings in degrees. Due to inaccuracies cause by the blending and alignment process, bearings were calculated to the nearest degree.

# CHAPTER 4 – PROTOTYPE RESULTS AND ACCURACY ASSESSMENT

## 4.1 Introduction

The software portion of the prototype developed for this research was designed and constructed in the C (GNU gcc compiler, Cygwin port) and Java (JDK1.2) programming languages for image processing and panoramic visualization/user interface development respectively. Both development environments are publicly available via the Internet at no cost (SUN, 2001; Cygwin, 2001). The system was tested primarily on a Pentium III 400 MHz processor with 258 Mb of RAM; however, the prototype is designed to be deployed on a standard PC-based platform with minimal additional software requirements (Java Virtual Engine, GNU DLLs)[3].

Throughout the development of the prototype, testing of specific system components was performed to determine the effectiveness of the proposed solution. Specifically, the brute force correlation matching technique was evaluated against the adaptive matching technique for the purposes of aligning and stitching a warped panoramic image sequence. Further, the non-calibrated "normal case" stereo rig setup was compared with the calibrated space intersection setup for the purposes of object to camera distance calculation accuracy. Finally, the computational efficiency of the

[3] It is worth noting that the Java interface can be deployed in any operating environment running the Java Virtual Machine.

prototype was considered. The results from these comparisons form the focus of this chapter.

### 4.2 Stitching (Image Matching) Implementation Evaluation

The two matching approaches were designed, implemented, and compared against a manually derived set of reference matches. In general, the approaches can be evaluated based on 1) the quality of the match, and; 2) the relative efficiency for finding a match. The former essentially relates to how much the proposed match deviates from the true match, while the latter refers to the speed at which a quality match can be found. In the context of panorama creation, the quick determination of quality matches is paramount since a mismatch detracts from the visual clarity of the output mosaic thereby reducing the value of the overall panoramic imaging system.

Twelve (12) image pairs corresponding to a test location (City Hall, Fredericton, New Brunswick) were processed using the brute force correlation and the adaptive matching software module developed by the author. These image pairs correspond to a typical urban scene with roadway, building, and vegetation features represented (Figure 3.6). Each software module requires as input the names of the image pairs to match up, the size of the template (7 x 7 pixels), as well as an *a priori* translation estimate for each image pair (-180,0). Each module provides a listing of the calculated translation in $x$ and $y$ as output.

For this testing approach, the input parameters for each matching approach were identical. The *a priori* estimate, as previously mentioned, serves to assist the algorithm in finding a match by reducing the total number of pixels requiring examination. While the

*a priori* estimate is not necessary, it does help in difficult to match up image pairs. The results of the matching process for both techniques are shown below (Table 4.1).

**Table 4.1. Translation errors in *x* and *y* for brute force and adaptive matching approaches given an a priori matching estimate of (–180,0), kernel size of 7 x 7 pixels, and 384 x 256 image size**

| | Manual Selection (pixels) | | Brute Force Correlation (pixels) | | Error (pixels) | | Adaptive Matching (pixels) | | Error (pixels) | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Image Pair** | **x** | **y** | **x** | **y** | **x** | **y** | **x** | **y** | **x** | **y** |
| **0–1** | -198 | -2 | -167 | 10 | -31 | 12 | -198 | -3 | 0 | -1 |
| **1–2** | -172 | 3 | -172 | 4 | 0 | 1 | -174 | 3 | 2 | 0 |
| **2-3** | -169 | -4 | -159 | 4 | -10 | 8 | -171 | -4 | 2 | 0 |
| **3-4** | -159 | -3 | -123 | 0 | -36 | 3 | -159 | -3 | 0 | 0 |
| **4-5** | -181 | 2 | -178 | 2 | -3 | 0 | -184 | 1 | 3 | -1 |
| **5-6** | -163 | 0 | -154 | -4 | -9 | -4 | -163 | -2 | 0 | -2 |
| **6-7** | -176 | -2 | -176 | 0 | 0 | 2 | -176 | -2 | 0 | 0 |
| **7-8** | -172 | 0 | -172 | -3 | 0 | -3 | -172 | 0 | 0 | 0 |
| **8-9** | -173 | -2 | -134 | 4 | -39 | 6 | -173 | -2 | 0 | 0 |
| **9-10** | -194 | -2 | -197 | -2 | 3 | 0 | -196 | -3 | 2 | -1 |
| **10-11** | -177 | -3 | -123 | 6 | -54 | 9 | -178 | -3 | 1 | 0 |
| **11-0** | -106 | -5 | -197 | 2 | 91 | 7 | -107 | -5 | 1 | 0 |
| | | | | **RMS** | 36.343 | 4.981 | | **RMS** | 1.084 | 0.669 |

Translation in Table 4.1 is presented according to the standard image processing convention; that is, the origin of an image's coordinate system is the upper left corner of the image matrix, with *x* values increasing in value from the left to the right, and *y* values increasing from the top to the bottom. Thus, for image pair 0-1 (corresponding to the first and second image in the panoramic sequence respectively), image 1 must be shifted 198 pixels to the left, and 2 pixels up to form a perfectly matched mosaic. Thus, for the purposes of panoramic image mosaic creation, each overlapping image pair manifests a consistent translation for a seamless match.

There are several key trends within the summary of the results in Table 4.1. Specifically, it is clear that the *x,y* translation varies greatly across image pairs. The largest translation appears for image pair 0-1 while the smallest translation occurs for image pair 11-1 (corresponding to the last and first image in the sequence respectively). It is suggested here that this is likely a result of human error in rotating the panoramic/stereo-bar rig in constant intervals. Although a precisely calibrated interval marker template was installed on the tripod face to facilitate a consistent rotation interval, it is likely that some misalignment did occur due to operator error.

Further, it is interesting to note the consistent negative *y* translation or drift across virtually all image pairs. This negative *y* translation accumulates to an 18 pixel shift "upwards" from the first to the last image in the panoramic sequence. While every effort was taken to ensure a leveled tripod setup, it is suggested that the sheer weight of the rotating stereo bar caused a slight but consistent pull on the tripod, thus taking it out of its original leveled state. Additional counter weights can be added to the shorter end of the bar in an attempt to counter-balance the longer portion of the stereo bar. However, despite the accumulated drift in the *y* direction and the inconsistent *x* translation across the image sequence, quality output results can still be obtained through the use of the software modules developed by the author.

It is clear from Table 4.1 that the adaptive matching approach is far superior to the brute force correlation approach in terms of the quality of matches that it returns (RMS in *x,y* of 1.084, 0.699 compared with 36.343, 4.981). In terms of processing time, the brute force correlation technique required 721 seconds in processing time (average of 60

seconds per image pair), while the adaptive matching approach required 117 seconds for completion (average of 9.75 seconds per image pair). The brute force correlation matching technique fails to provide reasonable translation estimates in more than half of the total image pairs examined. While it can be expected that the brute force correlation technique fails for homogeneous regions such as roadways and building facades, a thorough examination of the image pairs corresponding to the highest relative translation errors (image pairs 0-1, 3-4, 8-9, 10-11, 11-1) revealed that even in these homogeneous regions, some distinctive features can be observed. For example, referring to Figure 3.6 (images 10 & 11 in the left column), there are distinctive line features within the roadway that have the effect of breaking up the homogeneity of the roadway in general. Thus, in theory, the brute force correlation matcher should be able to find and subsequently match up these features. Unfortunately, the brute force correlation approach is based on the assumption that the maximum cross-correlation examined determines the most likely match. Thus, while the matcher may examine these distinctive features, a high correlation score is rarely achieved at these locations. In fact, the highest correlation score occurs most often for highly texture-less and consistent regions. The results of this research indicate that the underlying brute force correlation matching assumption often fails and does not provide reliable results using real imagery.

The adaptive matching approach, on the other hand, provides for better matching results due to 1) its ability to extract a subset of high contrast pixels in both images, and 2) its assumption that the best match occurs where there is the most support. Thus, even if a high correlation score is achieved, it may not necessarily be selected as the optimal

match. The results obtained for this research indicate that the support concept in combination with correlation matching can provide for effective matching results.

### 4.3 Distance Calculation Evaluation

In order to evaluate the accuracy of the prototype with respect to the calculation of object/camera distances, forty (40) separate distance calculations corresponding to 40 distinct measuring points were obtained from the panorama generated for the test site location (City Hall, Fredericton, New Brunswick). These points are distributed throughout the panorama and although not randomly generated, they are designed to reflect a variety of object types and distances. At each point, three measuring techniques were applied: 1) total station derived real-world distance; 2) system calculation using the panoramic viewer developed by the author under the assumption of normal case photogrammetry; and 3) system calculation using the panoramic viewer developed by the author with fully calibrated cameras and tripod setup (interior and exterior/relative orientation). The panoramic/stereo rig setup is problematic since it is unlikely that the assumption of "normal case" photogrammetry hold. However, it is a worthwhile experiment to test what level of accuracy can be achieved under this assumption since it provides for simpler data reduction.

The collection of real-world check distances was accomplished in early April, 2001 at the main test location (City Hall, Fredericton, New Brunswick) using a survey grade total station (electronic distance measurement device). At these same measuring points, distances were also computed (average of 3 repetitions of the user selected

location) using the operational software prototype under the normal case photogrammetry assumption. Finally, the cameras and stereo setup were calibrated according to the methodology developed previously. Once again, distances were computed (average of 3 repetitions of the user selected location) using the operational software prototype under the fully calibrated setup assumption.

The results of the camera calibration using the modified DLT technique are shown in Tables 4.2 and 4.3 for the left and right cameras respectively. As discussed in previously, five calibration images were acquired for each camera using the test field developed exclusively for camera calibration by Liu (1991) and adapted for this research by the author.

**Table 4.2. Camera calibration results (interior orientation parameters) for the left camera given an image size of 384 x 256 pixels**

| **Photo #** | **Principal distance (mm)** | **$x_o$ in mm (pixels)** | **$y_o$ in mm (pixels)** | $K_1$ | $K_2$ | $P_1$ | $P_2$ |
|---|---|---|---|---|---|---|---|
| **1** | 27.708 | 1.092 (11.6) | 0.701 (7.5) | 1.12E-04 | -4.31E-08 | 2.05E-05 | 2.30E-05 |
| **2** | 27.485 | 0.452 (4.8) | 0.312 (3.3) | 2.13E-05 | -4.02E-08 | -4.10E-05 | 3.20E-05 |
| **3** | 27.800 | 0.723 (7.7) | 0.736 (7.9) | 3.24E-04 | -3.14E-08 | -3.70E-05 | 5.36E-05 |
| **4** | 27.706 | 0.312 (3.3) | 0.324 (3.5) | 9.23E-04 | -2.34E-08 | -3.20E-05 | 3.53E-05 |
| **5** | 26.806 | 0.297 (3.2) | 0.823 (8.8) | 3.57E-04 | -3.00E-08 | -4.60E-06 | 7.24E-05 |
| | | | | | | | |
| **Mean** | 27.501 | 0.575 (6.1) | 0.579 (6.2) | 3.47E-04 | -3.36E-08 | -1.88E-05 | 4.33E-05 |
| **RMS** | 0.405 | 0.336 (3.6) | 0.243 (2.6) | 3.51E-04 | 7.99E-09 | 2.62E-05 | 1.97E-05 |

**Table 4.3. Camera calibration results (interior orientation parameters) for the right camera given an image size of 384 x 256 pixels**

| Photo # | Principal distance (mm) | $x_o$ in mm (pixels) | $y_o$ in mm (pixels) | $K_1$ | $K_2$ | $P_1$ | $P_2$ |
|---|---|---|---|---|---|---|---|
| **1** | 26.458 | 0.725 (7.7) | 0.163 (1.7) | 3.10E-05 | 1.92E-08 | 2.11E-05 | 1.53E-06 |
| **2** | 26.875 | -0.327 (-3.5) | 0.289 (3.1) | 2.87E-05 | -1.03E-08 | 2.10E-05 | 2.85E-06 |
| **3** | 26.093 | 0.256 (2.7) | 0.246 (2.6) | 3.18E-05 | -1.78E-08 | 2.17E-05 | 9.87E-05 |
| **4** | 27.013 | 0.453 (4.8) | 0.397 (4.2) | 1.89E-05 | -1.31E-08 | 3.08E-05 | 7.93E-05 |
| **5** | 26.324 | 0.343 (3.7) | 0.129 (1.4) | 2.46E-05 | -1.00E-08 | 3.45E-06 | 2.30E-06 |
| | | | | | | | |
| **Mean** | 26.553 | 0.290 (3.1) | 0.245 (2.6) | 2.70E-05 | -6.40E-09 | 1.96E-05 | 3.69E-05 |
| **RMS** | 0.383 | 0.387 (4.1) | 0.106 (1.1) | 5.32E-06 | 1.46E-08 | 9.94E-06 | 4.80E-05 |

As expected, the results of the camera calibration indicate that neither camera is very stable. This result is perhaps not surprising since non-metric cameras are well known to possess geometric instabilities that manifest in different calibration parameters per photograph. As suggested by Faig (1989), these types of instabilities must be taken into consideration when evaluating the accuracy of the generated results.

The mean values of the principal distance, principal point offsets, and distortion parameters of the 5 calibration images for the left and right cameras were used to determine the relative orientation parameters (Table 4.4) of each of the stereo pair from test location through the direct analytical calibration approach (dependent pair relative orientation). For each stereo pair, 30 point pairs (*xy* left, *xy* right) were manually selected in the overlapping coverage for each stereo pair and used as inputs in the enforcement of the coplanarity condition as required for relative orientation.

**Table 4.4. Relative orientation parameters for stereo panoramic image pairs**

| Image Pair | ω (radians) | φ (radians) | κ (radians) | $B_y$ (ratio to $b_x$) | $B_z$ (ratio to $b_x$) |
|---|---|---|---|---|---|
| **0–1** | -0.00275 | 0.00409 | -0.00138 | 0.00311 | 0.01644 |
| **1–2** | -0.00058 | 0.00789 | -0.00467 | -0.00926 | 0.03237 |
| **2-3** | -0.00188 | 0.00384 | -0.00417 | -0.00467 | 0.03378 |
| **3-4** | -0.00260 | 0.00317 | 0.00737 | -0.01806 | 0.02039 |
| **4-5** | -0.00275 | 0.00409 | -0.00138 | 0.00311 | 0.01644 |
| **5-6** | -0.00262 | 0.00357 | -0.00160 | 0.00236 | 0.01698 |
| **6-7** | -0.00165 | 0.01475 | -0.00135 | -0.00435 | 0.02920 |
| **7-8** | -0.00102 | 0.00475 | -0.00463 | -0.00684 | 0.02992 |
| **8-9** | -0.00058 | 0.00789 | -0.00467 | -0.00926 | 0.03237 |
| **9-10** | 0.00031 | 0.00762 | -0.00624 | -0.01310 | 0.03272 |
| **10-11** | 0.00042 | 0.01736 | -0.00429 | -0.01544 | 0.04223 |
| **11-0** | 0.00420 | 0.08529 | 0.00551 | -0.04148 | 0.10313 |
| | | | | | |
| **Mean** | -0.00096 | 0.01369 | -0.00179 | -0.00949 | 0.03383 |
| **RMS** | 0.00199 | 0.02300 | 0.00420 | 0.01228 | 0.02333 |

Table 4.4 illustrates the repeatability of the stereo rig setup since, in theory, the relative orientation parameters should not vary considerably between pairs due to the controlled nature of the stereo camera setup used in this research. In this sense, repeatability refers to the range of relative orientation parameter changes from one stereo pair to another. However, Table 4.4 clearly shows that the angular rotation elements do indeed vary somewhat throughout the process of acquiring complete 360 stereo panoramic coverage. It is interesting to note that under the assumption of normal case photogrammetry, values in Table 4.4 should approach 0.

The comparison of real-world, normal case, and calibrated case distance calculations is shown in Figure 4.1 where the *x*-axis refers to 40 distinct measuring points found in the panoramic image (Figure 4.2).

**Figure 4.1. Prototype system accuracy (accuracy over distance from left camera): normal and calibrated case**

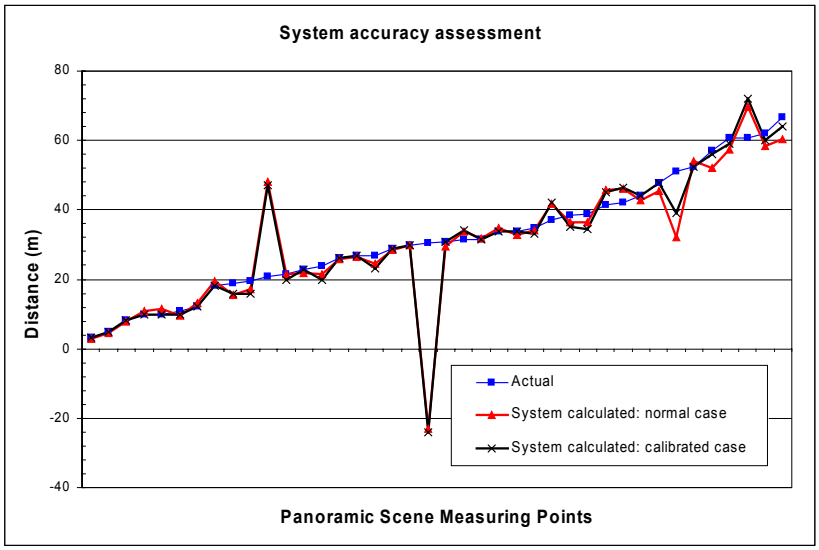
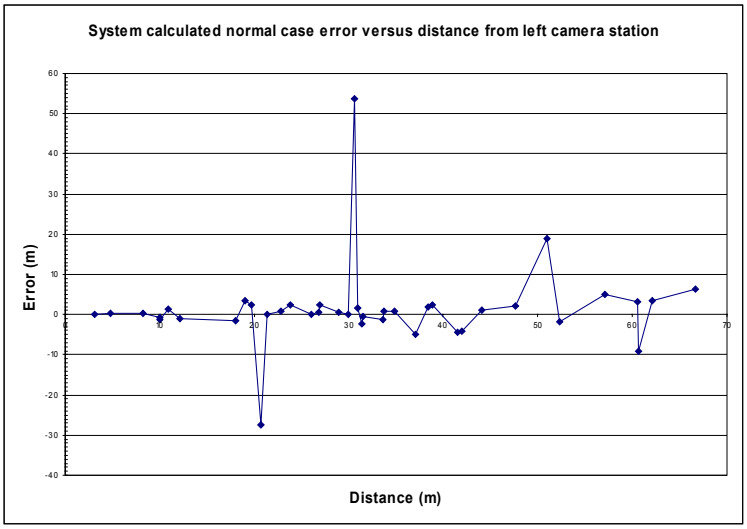



**Figure 4.2. Distribution of 40 manually selected panoramic scene measuring points**

The real-world distances and the system generated distances show a correlation of 0.832 for the normal case and 0.849 for the calibrated case. Further, the RMS error for the normal case is 10.37 metres, while for the calibrated case it is 10.19 metres. These errors exceed both the measuring accuracy of the prototype as well as the aforementioned objectives of this research. However, as can be seen from these results, blunders in the distance calculations greatly increase the overall inaccuracy. Four system-derived

distances were greater than 12 metres from their calibrated manually derived estimates, with a negative value being returned for check point location 14 (Figure 4.3).

**Figure 4.3. Absolute error versus distance for prototype system**




There appears to be no systematic over or under evaluation of the computer matched distances. On closer inspection of the input stereo pairs manifesting the poor distance estimates, it was revealed that poorly derived computer-matched distances were either in regions of homogenous pixel intensity, in areas experiencing temporal de-correlation, and occluded regions. Temporal de-correlation results from the non-simultaneous acquisition of the stereo pair, although every effort was made by the author to acquire images simultaneously[4]. Thus, although the scenes overlap, environmental

[4] An automatic exposure device linked to each camera would provide for effective simultaneous acquisition.

conditions changed and are not consistent between images (examples include: changing atmospheric conditions, or people and/or vehicles moving in and out of the scene). The occlusion problem is difficult if not impossible to solve. For example, referring to Figure 3.6, images 5 & 7, it is readily apparent that the street lamp is clearly visible in one scene yet not visible in the other. Thus, if a location is selected for these locations, the algorithm fails and it returns an erroneous match. Further, the assumption that the adaptive stereo match will find a high contrast pixel location in close enough proximity to approximate the original user selected pixel location may not always hold. This is perhaps not surprising since there is no guarantee that a distinctive pixel location will actually represent the original object selected.

In general, there are two factors that influence the accuracy of the system computed distances: 1) the quality of the computer generated stereo match, and 2) the geometric qualities of the cameras and stereo setup. It is interesting to note that the camera calibration approach used in this research fails to provide a distinct improvement in the system accuracy over the normal case assumption. This author suggests that this is not a failure of the calibration approach; rather, it provides support for the notion that the accuracy of the computer generated stereo match has the most significant impact on overall accuracy. Simply put, if the computer predicted stereo match is poor, it makes little difference if the camera is calibrated or not.

It is important to point out that check distances range from 1.5 metres to more than 65 metres in this accuracy assessment. While every attempt was made to include distances greater than 65 metres, it was observed that distances greater than 65 metres

could not be easy resolved in the images. Thus, in initial testing, the system derived distance estimates returned were very obviously inaccurate. Thus, image resolution plays a significant role in the accuracy of the computed distances. In general, closer objects were more accurately determined since more pixels in the image define the object.

Removal of the obvious blunders (> 12 metres) in the system generated distances yields a far more acceptable RMS error of 2.85 for the normal case and 2.67 for the calibrated case. For object distances less than 30 metres, these terms are reduced to 1.52 and 0.94 metres respectively. These results suggest that the non-calibrated stereo setup used in this research can provide adequate accuracy (within the +/- 1-3 metre threshold) for distances shorter than 60 metres from the stereo camera setup position. However, beyond 60 metres error values increase to well over acceptable thresholds (RMS > 10.0).

### 4.4 System Performance

The software developed for this research is currently completely automated and can find use with the non-specialist. Acquisition of imagery using the prototype stereo/panorama trip setup can be obtained in less than 10 minutes including setup, although longer acquisition periods can be expected for busy urban areas. Warping, matching, and seamless panoramic mosaic creation (blending) requires 139 seconds to process in batch processing mode using the imagery from the testing location (City Hall, Fredericton, New Brunswick). In terms of visualization, re-projection from the cylinder to the plane is achieved in near real time. Processing time for distance computations can

vary depending on the amount of time required to find a suitable match. In general, a distance can be returned after a user selection in 1-3 seconds.

# CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS

The thesis has outlined the design, development, and implementation of prototype for a new approach to virtual reality and GIS integration. A combined photogrammetric stereo and image processing approach was used to develop a set of hardware (stereo rig setup) and software (warping, matching, visualization, object/sensor distance determination) modules to provide proof of concept for this design. The virtual reality "immersion" effect presented as a scrolling 360-degree photo-realistic environment is convincing and effective.

The prototype described in this paper is a work in progress. Additional functionality such as real-time vector overlay was, unfortunately, beyond the scope of the project. It is anticipated that future research will be carried out to examine further enhancements to the basic design outline in this report. This author suggests that these enhancements must support the notion of ease of use and open concept design that were developed in this thesis.

However, the results of this testing show that, in general, ± 3 metre accuracy can be achieved for distance estimates under 60 metres using a completely un-calibrated and automated stereo and camera setup. This presents a high degree of simplicity in data reduction and computational efficiency since the cameras and stereo setup do not require extensive and labour intensive calibration. While the calibration process in itself is not extremely time-consuming, the calibration technique used in this research requires 3D

control information. In the context of the casual or inexperienced user, this type of information is simply unavailable.

While it is clear that camera calibration does present benefits with respect to improved system accuracy in previous research, the results from this research suggest that the camera calibration approach used here provides for only minimal accuracy improvement. This author suggests that this is not likely a result of poor calibration; rather, the quality of the stereo matching process has the greatest impact on the accuracy of the system. Thus, it is recommended that further research should focus on developing and refining superior stereo matching algorithms, and not necessarily camera calibration techniques.

It should also be highlighted that the pre-calibration approach used in this research amounts to an average of the calibration parameters for a given set of photographs. Thus, in order to achieve greater accuracy, a per photograph and per stereo pair calibration regime should be applied for each distance computed. It is recommended that future refinements to this prototype should explore the use of automatic self-calibration methodologies currently presented in the literature so as to take into account the instabilities of each photograph acquired.

Further, it is suggested here that currently available panoramic software processing tools are not adequate for the purposes of GIS integration. This unfortunately increases the development time and effort necessary for building a fully integrated GIS and PVR system. While the development of a comprehensive API was beyond the scope

of this research, developing such a framework would allow further research to focus on developing core applications rather than core functionality.

The novel image matching technique developed in the research for the purposes of image stitching can find relevance in a wide variety of image processing applications. It is recommended that the technique be extended to other types of image matching problems in different disciplines, such as remote sensing, digital aerial photography, and object recognition in order to assess whether the corner extraction methodology provides for effective matching results.

This thesis has demonstrated the value of the image-based rendering approach, namely panoramic virtual reality, in traditional planimetric map-based environments. While the road from idea, to design, and then to implementation is often long and challenging, the successful implementation of a working spatially enabled panoramic imaging prototype described in this thesis merits further research and development.

## REFERENCES


Abdel-Aziz, Y.I., and Karara, H.M. (1971). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry,* Urbana, Illinois: pp. 1 – 18.

Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3), pp. 283-310.

Apple. (2001). Apple Computer, Inc. http://www.apple.com

ASPRS. (1980). Manual of Photogrammetry. Ed. Slama, C. American Society of Photogrammetry and Remote Sensing. Falls Church, Virginia.

Barnard, S., and Thompson, W. (1980). Disparity analysis of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 2: pp. 333-340.

Beaudet, P.R. (1987). Rotationally invariant image operators. International Joint Conference on Pattern Recognition. pp. 579-583.

Bernard, M., Boutaleb, A., Kolbl, Q., Penis, C. (1986). Automatic Stereophotogrammetry: Implementation and Comparison of Classical Correlation Methods and Dynamic Programming Based Techniques. International Archives of Photogrammetry & Remote Sensing. Vol. 26-3/3.

Canny, J. (1983). Finding edges and lines in images. MIT technical report AI-TR-720.

Cygwin. (2001). http://sources.redhat.com/cygwin/

Chai, J. and De Ma, S. (1997). Robust epipolar geometry estimation using genetic algorithm. Pattern Recognition Letters, 19: pp. 829-838.

Chen, S.E. (1995). QuickTime VR – an image-based approach to virtual environment navigation. In*: Proceedings of SIGGRAPH '95 Computer Graphics Conference, ACM SIGGRAPH, August*, pp. 29-38.

Chapman, D., and Deacon, A. (1998). Panoramic imaging and virtual reality – filling the gaps between the lines. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53: pp. 311-319.

Cruz-Neira, C., Sandin, D.J., and DeFanti, T.A. (1993). Virtual reality: the design and implementation of the CAVE. In: *Proceedings of SIGGRAPH '93 Computer Graphics Conference, ACM SIGGRAPH, August,* pp. 135-142.

de La Losa, A., and Cervelle, B. (1999). 3D Topological modeling and visualization for 3D GIS. *Computers and Graphics*, 23 (4), 469-478

Derenyi, E. E. (1996). *Photogrammetry: the concepts*. Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada.

Dodge, M., Doyle, S., Smith, A., and Fleetwood, S. (1998). Towards the Virtual City: VR & Internet GIS for Urban Planning. http://www.casa.ucl.ac.uk/publications/birkbeck/vrcity.html

Drewniok, C., and Rohr, K., (1996). Automatic exterior orientation of aerial images in urban environments. *International Archives of Photogrammetry & Remote Sensing*, 31(3), pp. 146-152.

Dykes, J. (2000). An approach to virtual environments for visualization using linked geo-referenced panoramic imagery. *Computers, Environment, and Urban Systems*, 24(1): pp. 127-152.

El-Ansari, M., Masmoudi, L, and Radouane, L. (2000). A new region matching method for stereoscopic images. *Pattern Recognition Letters*, 21: pp. 283-294.

El-Hakim, S.F., Brenner, C., and Roth, G. (1998). A multi-sensor approach to creating virtual environments. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53: pp.379-391.

ERDAS. (2001). *ERDAS Virtual GIS*, ERDAS, Inc. Atlanta, GA.

ESRI. (2001). *ArcView 3D Analyst*. ESRI, Inc, Redlands, CA.

Faig, W. (1989). Non-metric and semi-metric cameras: data reduction. In: *Non-metric Photogrammety*, Ed: Karara, H.M. American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia.

Faugeras, O.D., Q.-T. Luong, and S.J. Maybank. *Camera self-calibration: theory and experiments*. In Proc. European Conference on Computer Vision, pages 321-334, SantaMargerita, Italy, 1992

Germs, R., Van Maren, G., Verbree, E., and Jansen, F.W. (1999). A multi-view VR interface for 3D GIS. *Computers & Graphics*, 23: pp. 497-506.

Gilles, S. (1996). Description and experimentation of image matching using mutual information. Technical Report: Oxford University. pp. 19.

Grimson, W. (1985). Computational experiments with a feature based stereo algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7(1): pp 17-34.

Harris, C. (1987). Determination of ego-motion from matched points. Proceedings of the Alvey Vision Conference (Cambridge).

Harris, C., and Stephens, M. 1988. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference (Manchester). pp. 147-151.

Hearnshaw, H.M., and Unwin, D. (1994). *Visualization in Geographical Information Systems*. Wiley, Chichester, pp. 243.

Heike, C. (1997). Automation of interior, relative, and absolute orientation. ISPRS Journal of Photogrammetry & Remote Sensing, 52: pp 1-19.

Heikkilä, J. & Silvén, O. (1997) A Four-step Camera Calibration Procedure with Implicit Image Correction. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico*. pp. 1106-1112.

Huang, Y.D. (1998). Capturing the third dimension by terrestrial photogrammetry. GIM, June: pp. 24-27.

Huang, B., and Lin, H. (1999). GeoVR: a web-based tool for virtual reality presentation for 2D GIS data. *Computers & Geosciences*, 25: pp. 1167-1175.

Huttenlocher, D.P., Noh, J.J., and Rucklidge, W.J. (1993). Tracking non-rigid objects in complex scenes. In: Proceedings of the Furth International Conference on Computer Vision, 10(3): pp. 93-101.

IPIX. (2001). Internet Pictures Corporation. http://www.ipix.com

Kalawsky, R.S. (1993). The Science of Virtual Reality. Readings: Addison-Wesley.

Kang, S.B. (1998). Geometrically valid pixel reprojection methods for novel view synthesis. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53: pp. 342-353.

Karara, H.M., and Abdel-Aziz, Y.I. (1974). *Photogrammetric potentials of non-metric cameras*, Civil Engineering Studies, Photogrammetry Series (No. 36), University of Illinois.

Kitchen, L., and Rosenfeld, A. (1982). Grey-level corner detection. Pattern Recognition Letters, 1: pp. 95-102.

Kodak. (2001). http://www.kodak.com

Koller, D., Daniilidis, K., and Nagel, H.-H. (1993). Model-based object tracking in monocular image sequences of road traffic scenes. International Journal of Computer Vision, 10(3): pp. 257-281.

Kruger, W., Bohn, C-A., Frohlich, B., Schuth, H., Strauss, W., and Wesche, G. (1995). The responsive workbench: a virtual work environment. In: *IEEE Computer Society IEEE Computer July*, 28(7): pp.42-48.

Lucas, B.D., and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th Imaging Understanding Workshop*. pp. 121-130.

Liu, X. (1991). *Calibration of non-metric stereo cameras*. Thesis: Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada.

Malmstrom, H. (1986). Measuring ground control for satellite image rectification. Schriftenreiche des Instititues fur Photogrammetrie der Universutat Stuttgart, 11: pp. 127-135.

Melen, T. (1994). Geometrical modelling and calibration of video cameras for underwatr navigation. PhD Thesis Dissertation. Trondheim, Norway.

Moniwa, H. (1980). The concept of photo-variant self-calibration and its application in block adjustment with bundles. *International Archives of Photogrammetry*, vol. 23, part B10, pp. 113-130.

Moravec, H. (1977). Towards automatic visual obstacle avoidance. Proceedings of the International Joint Conference on Artificial Intelligence, vol 8: p 584.

McMillan, L., and Bishop, G. (1995). Plenoptic modeling: an image-based rendering system. In: *Proceedings of SIGGRAPH '95 Computer Graphics Conference, ACM SIGGRAPH, August*, pp. 39-46.

Noble, J. (1988). Finding corners. Image Vision and Computing. pp. 121-128.

Parker, J.R. (1997). *Algorithms for image processing and computer vision*. John Wiley and Sons, Inc., Toronto.

Prewitt, J. (1970). Object enhancement and extraction. In: Lipkin and Rosenfeld, Eds., Picture processing and psychopictorics. Academic Press, New York: pp. 75-149.

Raper, J., and McCarthy, T. (1994). Virtually GIS: The new media arrive. In: Proceedings of the Association for Geographic Information Conference '94, Association for Geographic Information. pp. 18.1.1-6.

Raper, J., McCarthy, T., and Williams, N. (1999). Georeferenced four-dimensional virtual environments: principles and applications. *Computers, Environment, and Urban Systems*, 22(6): pp. 529-539.

Rhyne, T.M. (1997). Going virtual with geographic information and scientific visualization. *Computers & Geosciences*, 23(4): pp. 489-491.

Sun Micro Systems. (2001). JDK1.2 Programming Environment. http://www.sun.com

Szeliski, R., and Kang, S.B. (1995). Direct methods for visual scene reconstruction. In*: IEEE Workshop on Representation of Visual Scenes, Cambridge, MA, June*, pp. 26-33.

Szeliski, R., and Shum, H.Y. (1997). Creating full view panoramic image mosaics and environment map. In: *Proceedings of SIGGRAPH '97 Computer Graphics Conference, ACM SIGGRAPH, August*, pp. 251-258.

Unwin, D. (1997). The virtual field course. In: *Second International Conference on GIS in Higher Education, Columbia, MD, September 14,* pp. 4.

VRML. (1996). VRML 2.0 Specifications. http://www.vrml.org

Wolf, P. R. (1983). Elements of Photogrammetry. McGraw-Hill.

Woodfill, J., and Zabih, R.D. (1991). *An algorithm for real-time tracking of non-rigid objects.* In: Proceedings of the American Association for Artificial Intelligence Conference

Zhang, Z., Deriche, R., Faugeras, O., and Luong. Q-T. (1995). *A robust technique for matching two calibrated images through the recovery of the unknown epipolar geometry*. Artificial Intelligence. 78: pp.87-119.

# VITA

| | |
|---|---|
| Candidate's Full Name: | Stephen Rawlinson |
| Universities attended: | University of Guelph, Bachelor of Science (Environmental Sciences), 1997 |
| Publications: | Rawlinson, S., Y.C. Lee, and Y. Zhang (2001). Development of a software prototype for the geo-referencing and visualization of non-metric close range Photogrammetry in a GIS environment. *ASPRS 2001 Annual Conference,* St. Louis, Missouri, April 23-27, 2001<br><br>Maher, R., Rawlinson, S., and Thomas, V. (2001). Enabling Geographic Information Systems and Remote Sensing in a graduate curriculum for Natural Resources Management: a case study of the COGS-BIOTROP relationship, *Cartographica.*<br><br>Rawlinson, S., and Thomas, V. (1999). Integration of colour theory and digital elevation models and an alternative to traditional stereoscopic terrain visualization. *Proceedings of the 3rd Workshop of Electro-Communications and Information (WECI-III),* Institut Teknologi Bandung, Indonesia, March 3-4, Bandung, Indonesia. |
| Conferences Presentations: | Rawlinson, S., and Lee. Y.C. (2001). Panoramic Virtual Reality Integration, *DigitalEarth 2001,* Fredericton, NB.<br><br>Rawlinson, S., Lee,Y.C., and Zhang, Y. (2001). Development of a software prototype for the geo-referencing and visualization of non-metric terrestrial photography in a GIS environment, *ASPRS 2001 Annual Conference,* St. Louis, Missouri, April 23-27, 2001<br><br>Rawlinson, S., Thomas, V., and Maher. R. (1999). Development of a GIS and remote sensing curriculum in a developing country: an Indonesian experience. *International Cartographic Association conference*, Ottawa, Canada, August, 1999. |