A WEB-BASED 3D VISUALIZATION PROTOTYPE SYSTEM FOR HIGH-RESOLUTION SATELLITE COLOUR STEREO IMAGES

PINGPING XIE

April 2004

TECHNICAL REPORT NO. 223

A WEB-BASED 3D VISUALIZATION PROTOTYPE SYSTEM FOR HIGH-RESOLUTION SATELLITE COLOUR STEREO IMAGES

Pingping Xie

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

April 2004

PREFACE

This technical report is an unedited 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, April 2004. The research was supervised by Dr. Yun Zhang, and it was 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:

Xie, Pingping (2004). *A Web-based 3D Visualization Prototype System for High-resolution Satellite Colour Stereo Images.* M.Sc.E. thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 223, University of New Brunswick, Fredericton, New Brunswick, Canada, 118 pp.

Abstract

This thesis presents a Web-based 3D Visualization Prototype System (W3VPS) for high-resolution satellite stereo images. The W3VPS automatically generates 3D colour images using stereoscopic techniques for Web-based applications; it uses IKONOS images as source data, but is capable of using any other form of high-resolution satellite stereo images.

The W3VPS uses a well-established 2D affine orientation model for automatic re-sampling of epipolar images and then forms 3D anaglyph images. The W3VPS does not require rigorous orientation parameters, DTMs, or GCPs as inputs for 3D generation. A new image-matching algorithm is developed for the automatic matching of correspondences. It modifies Zhang's [1995] SM method and introduces a new algorithm for eliminating outliers that is based on the regression diagnostic approach. This matching algorithm successfully and efficiently finds a set of correspondences and eliminates the majority of reported correspondence outliers. For refinement of the 3D effect, a new method is developed based on the Quadtree technique. Finally, a one-tier Web client/server architecture is implemented for 3D Web visualization and provides ready access and exchange of files and full functionality to a dispersed user community.

Research results demonstrate the possibility of visualizing a natural environment in colour 3D using high-resolution satellite images and presenting this on the Web through a fast, cost-effective system. Further, the successful implementation of the W3VPS shows that the integration of photogrammetric principles with Web technologies is available for further development of Web-based satellite stereo image applications.

Acknowledgements

I would like to thank my supervisor, Dr.Yun Zhang, for initiating this research, and for providing helpful advice as well as the high-resolution pan-sharpened stereo images. Many thanks are due to Mr. Rob Lunn, GIS supervisor of the City of Fredericton, New Brunswick for providing the IKONOS multi-spectral and panchromatic images. As well, I would like to thank NSERC for providing a grant which supported my graduate studies.

My final thanks, as always, go to my family, Hui Li and Melody Li, for their support, encouragement, and consideration.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

3D	Three-dimensional
API	Application Programming Interface
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EOA	Eliminating Outlier Algorithm
GCP	Ground Control Points
GIS	Geographic Information System
GUI	Graphic User Interface
HTML	Hypertext Mark-up Language
HTTP	Hypertext Transport Protocol
JAI	Java Advanced Imaging
JPEG	Joint Photographic Experts Group
LMedS	Least Median of Squares
LSM	Least-square Matching
MS	Multi-spectral
Pan	panchromatic
RS	Remote Sensing
SM	Strength of Match
TCP/IP	Transmission Control Protocol/Internet Protocol
URL	Uniform Resource Locater
W3VPS	Web-based 3D Visualization Prototype System
WWW	World Wide Web
XML	eXtensible Mark-up Language

Chapter 1

Introduction

Three-dimensional (3D) visualization of aerial and satellite stereo images has been used for obtaining 3D information about the earth's surface for decades [Toutin, 2001]. In the fields of cartography, Geographic Information System (GIS), remote sensing (RS), and other geo-spatial applications, effective 3D visualization has remained a topic of active research. 3D visualization helps to enhance the analyst's ability to interpret two-dimensional imagery. Geo-scientists, environmental planners, engineers, mapping experts, and military personnel benefit from 3D information, in terms of an improved understanding of local environments and ability to interpret spatial details; furthermore, under certain conditions, 3D information may be crucial for mission success.

The World Wide Web (WWW) provides a powerful communication paradigm for image exchange and display based on multimediality and browsing, and their open architectural standards, which make all kinds of data and information available publicly and privately [Fraternali, 1999]. Several research studies demonstrated that the Web access to spatial data is more efficient [Abel et al., 1998; Doyle et al., 1998; Landes, 1999; Sugumaran et al., 2000; Morrison and Purves, 2002]. Therefore, 3D information distributed over the Web will facilitate decision-

making. However, through a review of the literature and a search of the Web, a system for automatic colour 3D generation and Web-based colour 3D visualization using high-resolution satellite colour stereo images has not been found by the author. In this thesis, a prototype system is presented that achieves a very effective Web-based 3D visualization using an anaglyph mode based on human binocular matching sensitivity.

Building a Web-based 3D Visualization Prototype System (W3VPS) involves several research areas and techniques: stereoscopy; techniques for producing terrain simulation in 3D with satellite stereo images; the potential of high-resolution satellite images for 3D; and, Web-based Internet and distributed programming. The next four sections briefly describe the respective research areas and techniques.

1.1 Stereoscopy

Stereoscopy is defined as the perception of depth through human binocular vision. The most important feature of binocular vision is that two images merge into one spatial impression by simultaneously channelling two separate left and right views to the respective eyes. Binocular disparity (parallax) and convergence are the two psychological cues for stereoscopy. Binocular disparity is an interocular difference between the positions of corresponding images in each retina. The degree of the disparity depends on the angle of convergence. Convergence is the ability to converge the two axes of the eyes. Binocular disparity is the most important cue for depth perception and reproduces the natural process of human binocular vision.

In established photogrammetric theory, stereoscopy is defined as “the science and art that deals with the use of images to produce a three-dimensional visual model with characteristics analogous to that of actual features viewed using true binocular vision” [Prade et al., 1966]. Stereoscopic viewing of images has been the most common method for obtaining 3D information of the earth’s surface over the last fifty years [Toutin, 2001]. The principle of stereoscopy is usefully applied to the interpretation of aerial photographs and satellite imagery and to the computation of terrain elevation. During this period, many different optical-mechanical, analytical, and digital 3D photogrammetric systems have been developed for both aerial and satellite stereo photogrammetry based on cues, binocular disparity and convergence [Toutin, 2001].

1.2 Existing Techniques with Satellite Stereoscopic Images

In the last few years, the generation of a digital elevation model (DEM) has become one of the most important studies in 3D information generation with satellite stereoscopic images [Al-Rousan et al., 1997; Petrie, 1999; Toutin, 2001]. Existing techniques for producing a simulation of terrain in 3D are based on either draping a 2D image over a DEM, or on the use of a DEM and an image through computing a trigonometric distortion (or shift) of the imagery displayed in one of the colour bands. These techniques are suitable for dynamically displaying 3D on a computer screen. However, it is a long and expensive process to derive a DEM that meets the accuracy requirements of user communities.

On the other hand, in many digital photogrammetric workstations and 3D satellite stereoscopic image systems, a 3D effect is achieved directly from a pair of stereoscopic images by employing anaglyphic or polarization techniques. Amongst anaglyphic techniques, the most common way is to display simultaneously or to print a pair of white/black images on a screen or on paper. The resulting 3D image is monocolor. Polarization techniques require an expensive polarization screen and polarizing glasses for viewing. The expensive hardware requirement makes it prohibitive as a Web application; in addition, this technique cannot provide a printed image on paper. Very few systems have used the stereoscopic technique and inexpensive complementary glasses to visualize colour 3D for high-resolution satellite stereo images and for Web applications. The author found this information through a review of the literature and a search of the Web.

1.3 Potential of High-resolution Satellite Images for 3D

The successful launches of high-resolution (up to 0.61m) satellites with high revisit (1-4 days) capability such as EROS, IKONOS, and Quickbird have marked the start of a new era of large scale and detailed 3D visualization and mapping of the earth's surface. The new generation of high-resolution imagery not only maintains dominant spectral advantages such as including near infrared radiation spectrum, but also shows a more important aspect: it provides strong geometric capabilities (high resolution, photogrammetric stereoscopic capability, and revisit rate) [Li, 1998]. 3D viewing and identification of 3D features with different details can be obtained from different high-resolution images (panchromatic: 0.61m and 1m, multi-spectral:

2.44m and 4m) in different stereo models (across track, adjacent track and along track) [Fritz, 1996]. Cook et al. [2001] have reported that geometric accuracy has met or exceeded the design specifications. Different 3D geometric correction methods to generate orthorectified products have been examined and they can produce high positioning accuracies [Toutin and Cheng, 2002; Toutin et al., 2002a; Toutin et al., 2002b]. Li [1998] also concluded that high-resolution satellite imagery could be used for the generation and updating of (US) national mapping products at a map scale of 1:24,000.

1.4 Web-based Internet and Distributed Programming

1.4.1Web-based Internet Technologies

The Internet is a huge collection of inter-connected networks. A network is the connection of computers and other devices, arranged to allow the sharing of resources [Sebesta, 2002]. On the Internet, the data transmitted between devices faithfully follow a protocol, which is a formal description of message formats and rules for exchanging messages between devices. The Transmission Control Protocol/Internet Protocol (TCP/IP) is a single low-level protocol that allows all devices to communicate with each other.

The World Wide Web, known simply as “the Web”, is a hypermedia-based system of Internet servers that use the Hypertext Transport Protocol (HTTP) to transfer specially formatted documents. The documents are formatted by mark-up languages such as Hypertext Mark-up Language (HTML) and eXtensible Mark-up Language (XML) that support links between documents. These documents are

accessed through Web browser, and are provided by Web server. Resources are identified on the Web through Uniform Resource Locators (URLs). The World Wide Web is increasing in popularity as a host to various kinds of applications. The Web makes available point-and–click browsers, enhanced multimedia capabilities of browsers and distributed multimedia applications in handling images, along with open architectural standards [Fraternali, 1999]. The ubiquity of these capabilities, and the fact that most Internet users are Web users, enable the full integration of digital images with the Web in everyday computing.

1.4.2 Distributed Programming

A distributed system denotes a network of computer systems that are dispersed in multiple locations, within which the activities of each component are coordinated and the resources of the system are shared. The advantages of a distributed system include resource sharing, improved performance as compared to standalone systems, and extensibility. Because the distributed system possesses a structure useful to modern organisations and exploits the pervasive nature of communication technologies, distributed systems have been widely used in different domains of applications such as e-business, e-government, and Web GIS. The Web is a typical example of a distributed system. Web servers, which perform the resource management role, are housed in various computers throughout the network. Each Web server possesses full control over its contents at all times and holds a wide range of documents and information on diverse topics.

The programming models of distributed systems may be classified as either based in distributing computing or parallel computing. In distributed computing, processing occurs in many different places such as possible servers, Web sites, personal computers, handheld devices, or other smart devices that are software-hardware constructions within a network. Distributed computing may be modelled using two different programming models: the client/server model and the object-based model. In terms of computing tasks, parallel computing is typically applied to the solution of large problems. It is done by dividing the problem into many small tasks which are distributed and executed on multiple processors on multiple distinct computers. Parallel computing usually requires more inter-processor communications than distributed computing.

1.5 Thesis Objective

Some of the 3D systems developed to date provide stereoscopic viewing of high-resolution satellite stereo panchromatic images. Web technologies are increasing in popularity for use in different kinds of applications such as e-commerce, e-government, and Web GIS. However, there has not yet been a rigorous investigation of the integration of photogrammetric principles with Web technologies to develop a Web-based satellite stereoscopic image application; nor have there been any reports on automatic 3D colour image generation and Web-based 3D interactive visualization using high-resolution pan-sharpened satellite stereo images. The objective of this research is to develop and test a complete methodology for the generation of natural colour 3D stereoscopic images using

photogrammetric principles in an anaglyph mode, as well as an effective Web-based 3D visualization package using Web technologies.

An automatic prototype system is constructed to generate colour 3D images and to interactively view natural colour 3D through Web browsers. To build the prototype system, it is needed to implement the following five aspects:

1) Automatic image-matching to evaluate correspondences for the high-resolution pan-sharpened stereo images;
2) Automatic generation of epipolar images as stereo image pairs for easier and sharper stereo viewing and forming an anaglyph image by superimposing the epipolar stereo image pair;
3) Building a Graphic User Interface (GUI) as a 3D viewer to display the 3D images on the local computer and simulate the Web 3D display;
4) Designing and testing the one-tier client-server architecture to implement a Web-based 3D visualization; and
5) Automatic parallax refining to improve 3D viewing effects for Web-based dynamic display.

Finally, the implemented system is tested and evaluated quantitatively. The key design considerations of the system are: 1) low cost; 2) simplicity and ease of operations; 3) high efficiency; and 4) robustness and reliability. The research outlined in this thesis presents an approach to Web-based 3D visualization at a low cost. It also demonstrates that it is possible to model a natural environment (such as an urban area) in Web-based 3D using the new high-resolution satellite image data.

This thesis is organized as follows:

- Chapter 1 (Introduction) provides an introduction to the rationale, objective, and scope of the research.
- Chapter 2 (Background) provides: 1) information about the ability of humans to utilize stereoscopic information under operational conditions; 2) an examination and explanation of the epipolar geometry of the perspective images and linear pushbroom images; 3) an overview of the most commonly used image-matching techniques for automatic satellite image-matching; and 4) an overview of the models and technologies of distributed computing for Web-based 3D interactive visualization.
- Chapter 3 (System Design and Development) focuses on the design, development, and implementation of the W3VPS developed by the author.
- Chapter 4 (Testing and Evaluation) describes the design and implementation of the testing methodology used to evaluate the performance of the W3VPS.
- Chapter 5 (Conclusions and Recommendations for Future Work) states the conclusions of the research, discusses its significance, and suggests recommendations and extensions to the W3VPS.

Chapter 2

Background

The Web-based 3D visualization prototype system developed in this thesis is based on theories and technologies of human stereopsis (human binocular matching sensitivity), the epipolar geometry of linear pushbroom images, image-matching techniques, and Web-based distributed programming technologies. The essential concepts of these theories or technologies are described in this chapter.

2.1 Human Stereopsis

A sound understanding of the human ability to utilize stereoscopic information is fundamental to the task of designing a stereoscopic display system. This section reviews the basic concepts that are relevant to both stereoscopic depth perception and to designing a stereoscopic display system [Patterson, 1992; Patterson, 1997; Wolf and Dewitt, 2000].

2.1.1 Depth Perception

There are ten cues that facilitate depth perception, four of which are physiological and six psychological. These ten cues contain the information necessary to allow humans to relate an object contained in a 2D image to 3D space. The four physiological cues are: accommodation, convergence, binocular disparity,

and motion parallax (for the details, see [Toutin and Vester, 2003]). The six psychological cues are: retinal image size, linear perspective, aerial perspective, overlapping, shade and shadows, and texture gradient (for the details, see [Toutin and Vester, 2003]). In the fields of remote sensing and photogrammetry, binocular disparity is the most important cue for stereo viewing of aerial photographs and visible light satellite imageries.

2.1.2 Viewing Methods in 3D

Until the 19th century, 3D information was obtained from a 2D image by means of psychological cues such as perspective, shade, and shadows [Toutin and Vester, 2003]. Over the last 200 years, the physiological cues of binocular disparity and convergence have been exploited by mechanical-optical and digital 3D imaging methods for more effective 3D viewing. In remote sensing and photogrammetry, the most commonly used 3D imaging methods of stereo viewing are stereoscopy, anaglyphs, and polarization.

Stereoscopy

A stereoscopic viewing system concentrates two eyes so that they view two respective images captured from two different viewpoints with the aid of optical devices such as stereoscopes; this method allows people to see in 3D by strengthening the physiological cues of binocular disparity and convergence [Toutin and Vester, 2003].

Anaglyphs

Anaglyphic viewing systems also depend on the physiological cues of binocular disparity and convergence, along with the aid of complementary colour filters, such as red and blue viewing glasses, to separate the left and right projections [Toutin and Vester, 2003]. Anaglyph images can be viewed in 3D either by being shown on a monitor or on colour prints. Because stereo viewing in anaglyph mode is an easy and inexpensive way to view 3D, it has become very popular in many areas such as microscopic techniques and geomatics. The 3D anaglyph method is among the stereo methods used in geomatics to interpret and to collect 3D information directly from imagery.

Polarization

Polarization viewing systems differ only slightly from anaglyphic viewing systems, in that light polarizing filters (e.g. polarizing glasses) are used to separate the left and right projections along with using a polarization screen. The polarization method is an expensive way to view 3D and it requires special hardware, therefore it is not suitable for Web-based 3D visualization.

2.1.3 Stereoscopic Basics

Stereo viewing is the perception of depth in such a way that it reproduces the natural process of stereovision. Stereo viewing is based on X parallax and the associated convergence angle (parallactic angle). Figure 2-1 illustrates the phenomenon of creating the 3D or stereoscopic impression of objects by viewing

identical images of the object, and it also shows the relation between the perception of the object's depth and the *X* parallax and convergence angle.

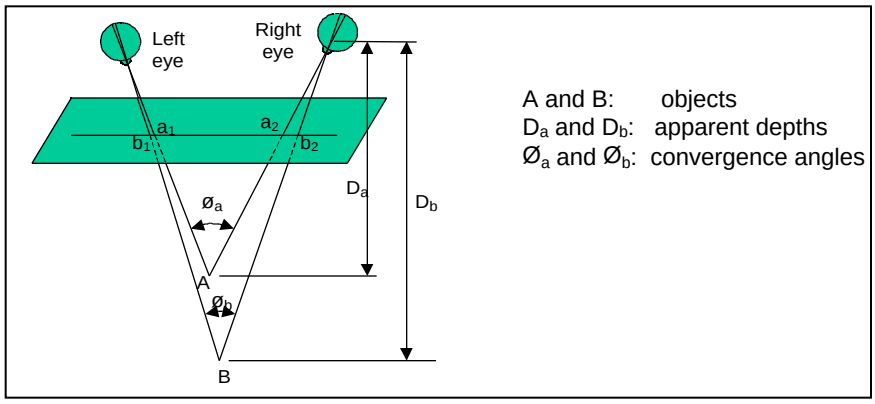

Figure 2-1. The depths of the object A and B, and convergence angles (after Wolf and Dewitt [2000])

X parallax

The *X* parallax (*X* disparity) is caused by a shift in the position of observation [Wolf and Dewitt, 2000; Toutin and Vester, 2003]. For generating satellite image stereo pairs, satellites collect image data at two different times, two different look angles, or two different beam positions. These changes cause the apparent displacement of the position of the object with respect to a certain frame of reference [Wolf and Dewitt, 2000; Toutin and Vester, 2003]. Figure 2-2 demonstrates two important characteristics of *X* parallax [Wolf and Dewitt, 2000; Toutin and Vester, 2003]: (1) the parallax of a point is directly related to the elevation of that point; and (2) the parallax is greater for high points than for low points.

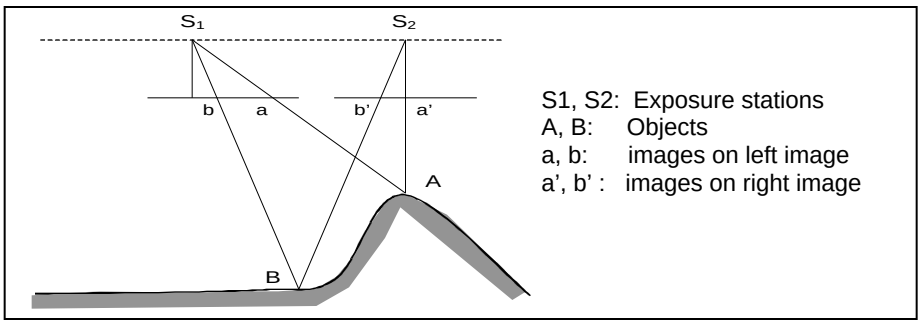

Figure 2-2. An example of the *X* parallax of vertical aerial photographs (after Wolf and Dewitt [2000])

Parallactic angle

The parallactic angle is formed by the intersection of the optical axes of two eyes converging on a certain point with binocular vision (see Figure 2-1). The closer the object is to the eyes, the larger parallactic angle it forms. The ability of humans to detect changes in the parallactic angle enables depth perception.

Y parallax

The *Y* parallax (*Y* disparity) is the difference in perpendicular distance between two images of an object from the air base line. The possible sources of *Y* parallax include improper orientation of the photos, variation in flight heights, and tilt of the photos. Research results [Stevenson and Schor, 1997] show that the human visual system takes advantage of the epipolar constraint, but that it is not restricted to it as an operational mode of perceiving 3D. However, for correct depth judgements and for clear and comfortable stereo viewing, the *Y* parallax should be minimized.

2.1.4 Stereoscopic Depth Perception

As discussed in section 2.1.3, stereoscopic viewing is induced by *X* parallax. This section discusses issues related to the *X* parallax in the design of a stereo display system, which creates a consistent and easily perceived impression of depth.

The Horopter

As shown in Figure 2-3, point F is the point on which the two eyes fixate; thus, the imaginary longitudinal horopter is the curved line passing through F [Patterson, 1992; Patterson, 1997]. Images of objects falling on the horopter stimulate the corresponding retinal points in the two eyes. Objects have zero disparities relative to F (for example object A in Figure 2-3). Those objects that are not on the horopter have non-corresponding points and non-zero disparity related to F (for example object B in Figure 2-3). The larger the depth in front of, or behind, the horopter, the larger the disparity in the perceived position of objects.

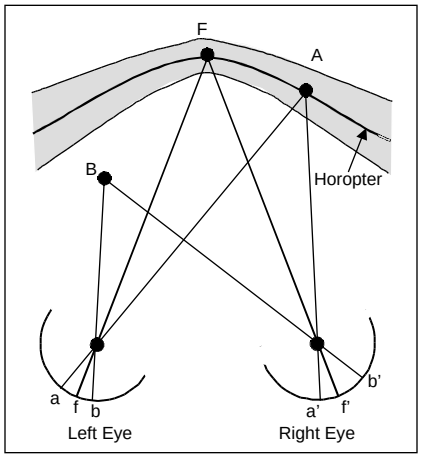

Figure 2-3. Top-down view depicting left and right eyes fixating on point F, the longitudinal horopter, stimuli A and B (from Patterson [1992])

Crossed versus Uncrossed Disparity

There are two ways to define disparity directions. The first definition is given in relation to the horopter. The term “crossed” denotes the disparity in front of the horopter, while “uncrossed” denotes the disparity behind the horopter. The second definition is related to certain applications such as the perception of a display device. Display screens are considered as the baseline of zero depth. The disparity related to the depth in front of screen is crossed; otherwise it is uncrossed (for the details, see [Patterson, 1992]).

In stereo viewing systems, the system design takes account of the effect of disparity directions because, in nature, humans perceive depth information separately in two directions, and one of the two directions is impaired for stereo-anomalous observers who are able to only perceive transient depth to a wide range of disparities presented in one of the depth directions. A thematic separation of depth perceptions in the crossed direction from those in the uncrossed direction is recommended in both psychophysical and physiological literature [Patterson, 1992; Patterson, 1997].

Binocular Fusion and Rivalry

Binocular fusion is the sensory process that blends the two corresponding monocular images into one perceived image [Patterson, 1992]. When the objects’ positions are far away from the horopter, images are diplopic. This leads to the process of binocular rivalry. The rivalry occurs as the visual system attempts to fuse the dissimilar images. Where there is a large disparity between corresponding

monocular images of an object, the rivalry is more pronounced (for the details, see [Patterson, 1992]).

During natural stereo viewing, both fusion and rivalry operate. When objects are close to the horopter (such as the grey area in Figure 2-3), images of objects are fused; otherwise, the images of objects are rivalrous.

Stereoscopic Displays

Depth perception in stereo viewing is produced by simultaneously presenting two separate left and right views to the respective eyes of an observer with a lateral shift between the positions of corresponding images. The quantitative prediction of depth perceptions can be derived from the geometry of stereopsis [Patterson, 1992; Patterson, 1997]:

$$d = S \quad D/(I + S) \qquad \text{for crossed disparity} \qquad 2\text{-}1$$

$$d = S \quad D/(I - S) \qquad \text{for uncrossed disparity} \qquad 2\text{-}2$$

where d is the predicted depth; S is the separation between the half-image of the display and is in proportion to the disparity; D is the viewing distance; and I is the interpupillary distance. There is an asymmetry between the crossed and uncrossed disparity directions, in which the depth with uncrossed disparity is greater than that with crossed disparity, given the same separation. Research results also show that the depth with crossed disparity follows predictions more closely than the depth with uncrossed disparity in most cases [Patterson, 1992; Patterson, 1997].

2.2 Epipolar Geometry of Linear Pushbroom Satellite Images

For easier, sharper, and more comfortable stereo viewing, epipolar images as stereo images are required to reduce the amount of Y parallax between stereo images. Given the parameters of the epipolar geometry of the stereo images, image processors can utilize this geometry for epipolar image re-samplings; stereo-matching algorithms, which determine the corresponding objects of stereo images, can also use this constraint to reduce the matching search spaces.

Applications of epipolar geometry are different for each image obtained. For example, close-range images are different from the satellite images because the epipolar geometry relies on the type of sensor that is used. Traditionally, in the computer vision community, an epipolar geometry based on the fundamental matrix has been used for close-range images [Zhang et al., 1995]. By comparison, in the photogrammetry community, an epipolar geometry based on the collinearity equations is used for aerial photographs [Wolf and Dewitt, 2000]. Although the epipolar geometry of linear pushbroom images is different from that of perspective images, an epipolar geometry clue can also be useful for processing satellite images obtained from linear pushbroom sensors [Kim, 2000]. In this section, the epipolar geometry of perspective images and its properties are explained; following this, the epipolar geometry of linear pushbroom images and its properties are presented.

2.2.1 The Epipolar Geometry

Figure 2-4 illustrates the property of epipolar geometry that a point in the left (or right) image corresponds to a unique curve in the right (or left) image, as defined

by the orientation and position of both left and right cameras and the sensor models. For example, if a hypothetical light beam (the dotted line in Figure 2-4) is emitted from the left exposure station S through the point q(x_l, y_l), then each point lying on this beam has a unique mapped point on the right image. Those mapped points describe a curved line, the epipolar curve of point q, on the right image. The conjugate point q' of q for the ground point Q(X, Y, Z) always lies on this epipolar curve.

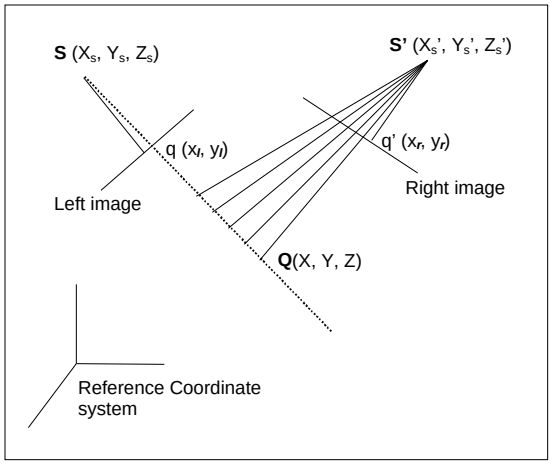

Figure 2-4. A general scheme of epipolar geometry (from Kim [2000])

2.2.2 The Epipolar Geometry of Perspective Images

Figure 2-5 illustrates the characteristics of the epipolar geometry of perspective images. The epipolar plane of object *A* is defined by triangle L_1L_2A. The epipolar curve lines are defined as the resultant lines of the intersection of the epipolar plane with the left and right planes. Because the intersection of any two planes is always a straight line, it is categorical that the epipolar lines of perspective

images are straight lines. These two epipolar lines form an epipolar line pair, which has the property that the correspondences of any points lying on one epipolar line in one image are always lying on another epipolar line in another image. This property markedly reduces the search spaces for the image match problem.

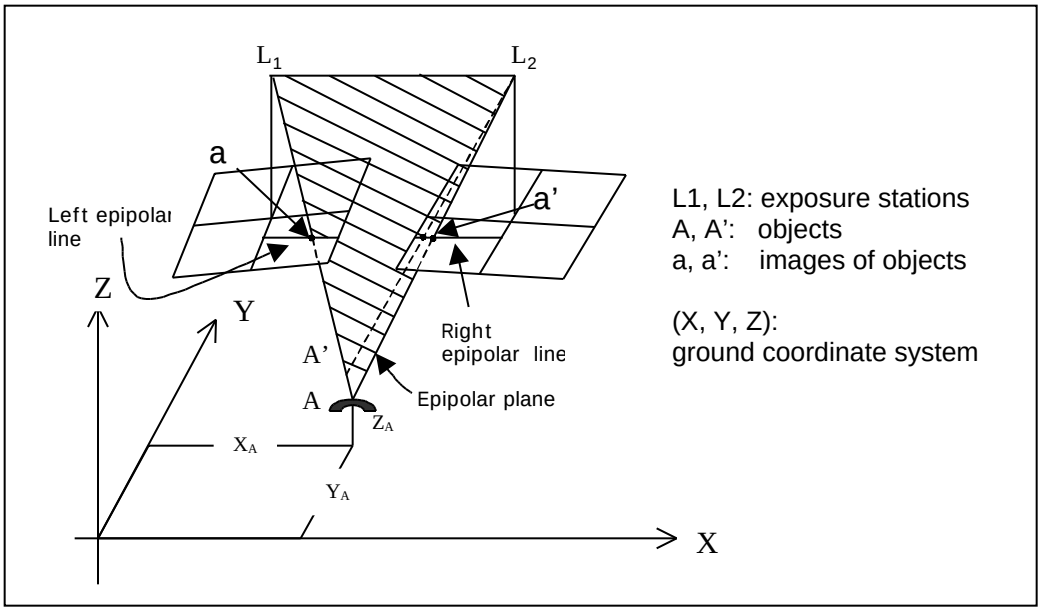

Figure 2-5. Epipolar geometry in a perspective view (after Wolf and Dewitt [2000])

Fundamental Matrix

Equation 2-3 defines an epipolar line of a point of the left image in the right image by using the fundamental matrix [Zhang et al., 1995; Kim, 2000; Forsyth, 2003]:

$$x_r(F_{11}x_l + F_{12}y_l + F_{13}) + y_r(F_{21}x_l + F_{12}y_l + F_{23}) + (F_{31}x_l + F_{32}y_l + F_{33}) = 0 \qquad 2\text{-}3$$

where (x_l, y_l) is the coordinate of the point q in the left image; and (x_r, y_r) is the coordinate of the conjugate point q' of the point q in the right image. F_{11} to F_{33} are

the coefficients of the fundamental matrix *F*. This equation also shows the property of epipolar lines of the perspective images; that is, that the epipolar lines are straight lines.

Collinearity Equations and the Coplanarity Condition Equation

In analytical phototgrammetry, the collinearity condition is the most useful and fundamental relationship. The condition is expressed by the collinearity of three points: exposure station, object, and its photo image, in a 3D space; see Figure 2-5, L_1, a, and A are lying a straight line. This condition is defined by two equations (equations 2-4 and 2-5): one equation for the *x* coordinate, and another for the *y* coordinate, in the image plane [Wolf and Dewitt, 2000].

$$x = x_0 - f\frac{\left[m_{11}(X - X_L) + m_{12}(Y - Y_L) + m_{13}(Z - Z_L)\right]}{\left[m_{31}(X - X_L) + m_{32}(Y - Y_L) + m_{33}(Z - Z_L)\right]} \qquad 2\text{-}4$$

$$y = y_0 - f\frac{\left[m_{21}(X - X_L) + m_{22}(Y - Y_L) + m_{23}(Z - Z_L)\right]}{\left[m_{31}(X - X_L) + m_{32}(Y - Y_L) + m_{33}(Z - Z_L)\right]} \qquad 2\text{-}5$$

where *m*'s are the functions of the rotation angles omega, phi, and kappa to the object space coordinate system; and *f* is the focal length of the camera. The *x* and *y* are the photo coordinates of the image point; and x_0 and y_0 are the photo coordinates of the principal point that usually is known from camera calibration; *X*, *Y*, and *Z* are object space coordinates of the object point; and X_L, Y_L, and Z_L are object space coordinates of the exposure station. The rotation angles omega, phi, and kappa define a sequence of three elementary rotations around the *x*, *y*, and *z* axes,

respectively, where the *xyz* coordinate system is parallel to the *XYZ* object system and is constructed with its origin at the exposure station.

The coplanarity condition holds that the two exposure stations, the object, and its image points in two image planes are in the epipolar plane (Figure 2-5). Because the expression of the coplanarity equation is difficult to calculate, it is not used as widely as the collinearity condition in analytical photogrammetry.

2.2.3 The Epipolar Geometry of Linear Pushbroom Images

Early in the research for the epipolar geometric relationship of linear pushbroom images, the exact epipolar curve was not easy to derive because of the complexity of the relationship and the difficulty in expressing it in mathematical form [Otto, 1998]. Subsequently, Kim [2000] derived an exact mathematical equation for the epipolar curve of linear pushbroom images using a collinearity-based sensor model.

The epipolar geometry shown in Figure 2-5 does not hold for linear pushbroom images any more because of the fundamental difference between perspective images and linear pushbroom images - the difference being that perspective images have only one fixed centre of perspective, while linear pushbroom images have one centre for each line. However, collinearity equations can be used for linear pushbroom images by modifying the equations for perspective images in the scanning direction [Kim, 2000].

In Figure 2-4, the relationship among exposure station S(X_S, Y_S, Z_S), q(x_l, y_l), and Q(X, Y, Z) can be expressed in a ground-referenced coordinate system with the parameterized function of k, as shown below:

$$(X, Y, Z) = (X_s, Y_{s,} Z_s) + k \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ -f \end{bmatrix} \qquad 2\text{-}6$$

where r's represent the functions of the rotation angles omega, phi, and kappa of the left image to the ground reference coordinate system; and f is the focal length of the camera. The collinearity equation of the conjugate point q'(x_r, y_r) of q for the ground point Q(X, Y, Z) in the right image can be expressed by the following equations:

$$0 = r'_{11}(X - X'_S) + r'_{21}(Y - Y'_S) + r'_{31}(Z - Z'_S) \qquad 2\text{-}7$$

$$y_r = -f \frac{r'_{12}(X - X'_S) + r'_{22}(Y - Y'_S) + r'_{32}(Z - Z'_S)}{r'_{13}(X - X'_S) + r'_{23}(Y - Y'_S) + r'_{33}(Z - Z'_S)} \qquad 2\text{-}8$$

where r's are the functions of the rotation angles omega, phi, and kappa of the right image to the ground-referenced coordinate system. By substituting (X, Y, Z) of equation 2-6 for (X, Y, Z) of equations 2-7 and 2-8, the equation of the epipolar curve is established in the following form [Kim, 2000]:

$$y_r = \frac{A_1 x_l + A_2 y_l + A_3}{(A_4 x_l + A_5 y_l + A_6)\sin Q(x_r) + (A_7 x_l + A_8 y_l + A_9)\cos Q(x_r)} \qquad 2\text{-}9$$

where A's are constants for a given scan line x_1; and $Q(x_r)$ is a quadratic function of x_r associated with the yaw angle variation of the right sensor. Equation 2-9 clearly shows that the epipolar curve of a point is not a straight line but a hyperbola-like shape. Kim [2000] also concluded that the epipolar curve pair does not, strictly speaking, exist for linear pushbroom images but exists only "locally". Figure 2-6

illustrates that the epipolar curve is not linear but may be approximated by a straight line for a small range.

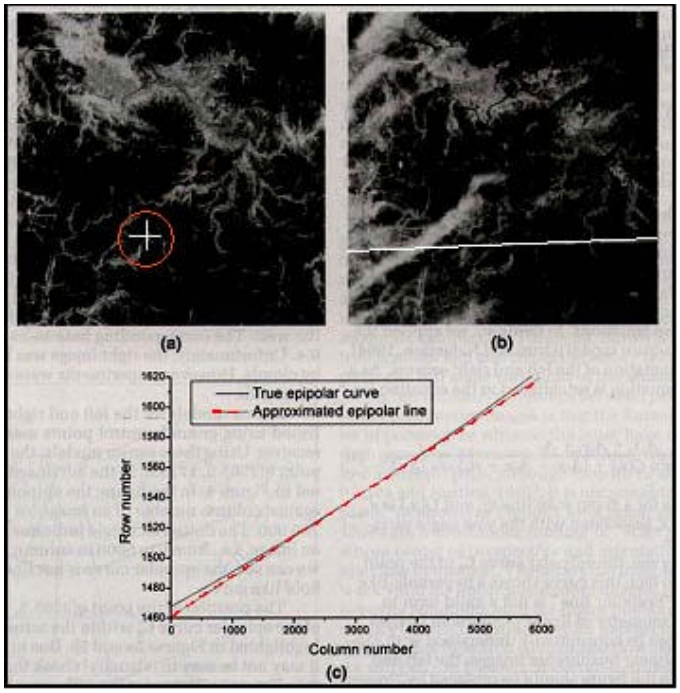

Figure 2-6. A true epipolar curve and an approximated epipolar line for a spot stereo pair. (a) the position of point q(2565.5, 1715.8), (b) an epipolar curve, and (c) difference between true epipolar curve and its approximation (from Kim [2000]).

2.3 Correspondence Problem

In both computer vision and photogrammetry, the correspondence problem concerns the problem of image-matching. Image-matching determines where correspondences (corresponding pixels, points, or other features, such as edges and regions) occur in images taken from stereo cameras for the same entity in the 3D world. It is also a major problem for colour 3D image generation in this research.

Matching techniques can be categorized broadly into area-based and feature-based image-matching, or a combination thereof. Area-based techniques are based on matching texture regions in two images, while feature-based techniques rely on the characteristics of abstract features. Both techniques are combined to deal with the types of images associated with specific applications [Cochran and Medioni, 1992; Weng et al., 1992; Zhang et al., 1995; Zhou and Shi, 2002].

Ambiguities always exist during image-matching processes for the determination of correspondences in stereo images because of factors such as noise, lighting variation, occlusion, and perspective distortion. For example, a point in the first image may correspond to several points in the second image (called candidate matches), and vice versa. Normally, some constraints and assumptions are applied to reduce the ambiguities and to speed up computations. Before the matching techniques are studied, the next section outlines some assumptions and constraints which illustrate the inherent complexity of the corresponding problem.

2.3.1 Assumptions and Constraints

In general, the constraints and assumptions can be classified into two broad categories based on models of image generation and properties of scene objects [Klette et al., 1998]:

(1) Based on models of image generation (geometric constraints):

- The location of the epipolar lines (epipolar constraint)
- The uniqueness of the assignment (uniqueness constraint)

- The compatibility of image intensity values (photometric compatibility constraint)
- The geometrical similarity of certain features (geometric similarity constraint)

(2) Based on properties of scene objects:

- The order of the projected points in the image (ordering constraint)
- The continuity of the disparities (continuity constraint)
- The continuity along the orientation edges (figural continuity constraint)
- The compatibility of certain features (compatibility constraint)
- Disparity limit or the disparity gradient limit

Generally, not all of these constraints are considered simultaneously by the operation of a single algorithm of the image-matching. The following sections give an overview of the constraints and assumptions applied to the design of correspondence analysis processes in this research.

2.3.1.1 Epipolar Line Constraint

The epipolar line constraint is based on a geometrical law which is based on the stereo image's epipolar geometry described in section 2.2.1: as illustrated in Figure 2-4, the conjugate point q' of a point q in the left image for object Q can only lie on the corresponding epipolar line in the right image which is uniquely defined by q. Therefore, to search for the corresponding point in the right (or left) image of a point in the left (or right image), the search space can be restricted to those points

lying on the corresponding epipolar line. This simplifies the search task for a corresponding pair from a two-dimensional problem to a one-dimensional one.

In general, the determination of epipolar lines requires either a geometrically calibrated image acquisition set-up or a set of corresponding pairs.

2.3.1.2 Uniqueness, Compatibility, and Similarity

These assumptions depend on the shape, illumination, and distances of the scene objects, whereas the epipolar constraint depends on the geometrical law.

Uniqueness assumption: Typically each point in one image of a stereo pair has one exactly corresponding point in another image. However, there are exceptions. For example, in a 3D space, two objects lying on a same light beam project into one pixel in one image but project into two different pixels in another image. This particular type of exception is not considered further in this research.

Compatibility assumption for intensity values: if two points of a stereo pair can correspond to each other, the intensity difference between them should be below a selected threshold value. Because a single pixel is sensitive to noise, its intensity value is not used to establish the correspondence. In practice, the intensity values in the local neighbourhood, generalized over a range of surrounding pixels, are used for the correspondence analysis.

Geometric similarity: Normally, two types of assumptions are made, described as follows: (1) Angle criterion - suppose there is a point pair (m, n) of a stereo pair; if the angle between m and n is less than a specified threshold, they are considered as a corresponding pair; and (2) Length criterion - suppose there is a line

segment pair (L1, L2) of a stereo pair; if the length difference between L1 and L2 is less than a threshold, they are considered to be a corresponding line segment pair.

2.3.1.3 Continuity of Disparities

This assumption, that the disparities of corresponding pairs continuously change everywhere in a stereo image pair, is based on the assumption that object surfaces are opaque and smooth.

Continuity assumption for disparity: Suppose that there is a corresponding pair $P_{L1}(x_{L1}, y_{L1})$ and $P_{R2}(x_{R1}, y_{R1})$ for a given stereo image pair. One point $P_{L2}(x_{L2}, y_{L2})$ in one image of the stereo pair is a near neighbourhood of PL1. The correspondence of P_{L2A} in another image can only be such a point P_{R2} (x_{R2}, y_{R2}) that the disparity difference between the first pair (P_{L1}, P_{R1}) and the second pair (P_{L2}, P_{R2}) is less than a certain threshold (θ) [Klette et al., 1998]:

$$\left|\sqrt{(x_{L1}-x_{R1})^2+(y_{L1}-y_{R1})^2}-\sqrt{(x_{L2}-x_{R2})^2+(y_{L2}-y_{R2})^2}\right|<\theta \qquad 2-10$$

$$\left|\,|x_{L1}-x_{R1}|-|x_{L2}-x_{R2}|\,\right|<\theta \qquad 2\text{-}11$$

Exceptions occur when surface points are along the orientation edge. Typically, this continuity constraint is applied within small image regions instead of within the whole image.

2.3.2 Area-based Matching

As described in Zhang et al. [1995], Do et al. [1998], Klette et al. [1998], Schenk [1999], and Zhou and Shi [2002], area-based matching rests on the

assumption that corresponding pairs have a similar intensity value. Area-based techniques have been successfully applied to aerial and satellite images [O'Neill and Denos, 1996; Jokinen, 1998; Baillard and Dissard, 2000]. However, because of this underlying assumption, they are applicable to stereo images that have good textures and small differences in the intensity values and tend to break down where there are occlusions and featureless regions. Figure 2-7 illustrates the concept of area-based matching. The objective of area-based matching is to compare the intensity value distribution of several neighbouring pixels between a template and a matching window. This comparison is done using a similarity measure defined for the intensive values in the image window (template and matching window). The best-known area-based matching methods are cross-correlation and the least-square matching.

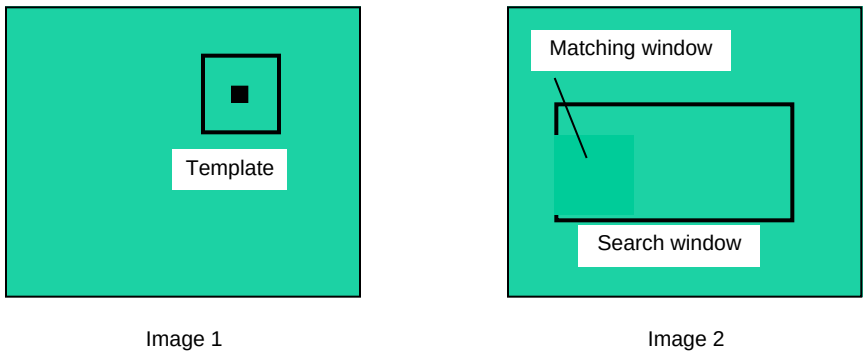

Figure 2-7. The concept of area-based matching (from Schenk [1999]). Template: image patch fixing a position in one of the images. Search window: the search space in the other image. Matching window: image patch compared with the template in the other image

Cross-correlation Matching Technique

There was a lot of research done on digital correlation techniques for image-matching from the early 1970's to the mid 1980's. Since Kreiling (1976) and other pioneers used digital correlation for generating DEMs automatically in early 1970's, correlation techniques have become accepted for finding conjugate points in photogrammetry. The cross-correlation factor is used, with the normalized cross-correlation coefficient between template and matching windows being defined as:

$$\rho = \frac{\sigma(I_1 I_2)}{\sigma(I_1) \times \sigma(I_2)}, \quad -1 \le \rho \le 1 \qquad 2\text{-}12$$

$\sigma(I_1 I_2)$ covariance of image patches I_1 and I_2

$\sigma(I_1)$ standard deviation of image patch I_1

$\sigma(I_2)$ standard deviation of image patch I_2

where,

$$\overline{g_{I_k}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} g_{I_k}}{n \times m}, \ \overline{g_{I_k}} \text{ the mean of image patches } I_{k,}\ k = 1,2 \qquad 2\text{-}13$$

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(g_{I_k}(x_i, y_j) - \overline{g_{I_k}}\right)}{n \times m - 1}}, \qquad 2\text{-}14$$

$g_{I_k}(x_i, y_j)$ are image functions of image patches $I_k, k = 1,2$

$$\sigma(I_1 I_2) = \frac{\sum_{i-1}^{n}\sum_{j=1}^{m}\left(\left(g_{I_1}(x_i, y_j) - \overline{g_{I_1}}\right)\left(g_{I_2}(x_i, y_j) - \overline{g_{I_2}}\right)\right)}{n \times m - 1} \qquad 2\text{-}15$$

An optimal match manifests a normalized cross-correlation value of 1 since, in theory, the stronger the correlation between two correlation windows, and the greater the likelihood that they are similar. However, the maximum is not determined very well in some real cases so that the reliability of the conjugate point is low.

Least-square Matching Technique

As in the cross correlation matching technique, Least-square matching (LSM) is also based on the similarity of grey levels. The objective of LSM is to minimize grey level differences between the template and the matching window by changing the matching window's position and shape, as determined in the adjustment process. The change process stops when the grey level differences between the deformed matching window and the template reach a minimum. This method not only considers illumination and reflectance differences between the two images but also the geometric distortion of the regular image tessellation caused by unknown orientation parameters, tilted surface patch, a surface patch with relief, etc. All pixels in the final deformed matching window become conjugate with their counterparts in the template, and thus LSM is regarded as the most accurate image-matching technique. However, it is very sensitive with respect to the quality of the approximations, requiring, for example, the prior information of correspondences within an accuracy of a few pixels. Thus, LSM always follows the application of other matching techniques such as cross correlation and, as a final step, improves the match accuracy.

Issues

Some issues must be addressed independently from the area-based matching methods. A brief overview is given below.

Location of template: Under certain conditions, the locations of the template will cause area-based matching to fail - where, for example, the template is placed on an area that is occluded in the other image.

Size of template: Finding the optimal size of the template is important. As size increases, the significance of ambiguities in the matching operation usually decreases but the geometric distortion increases. Unfortunately, there is no accepted standard for the optimal size of the template, and so a compromise such as trying different template sizes must be found.

Size and location of search window: The size of the search window is not critical, but initial selection of an appropriate size can reduce the search space to speed up the matching algorithm. The location of the search window, however, is crucial.

2.3.3 Feature-based Matching

As discussed in Zhang et al. [1995], Klette et al. [1998], and Schenk [1999], feature-based matching relies on comparisons of image features (corners, edges and regions) extracted from the original intensive image. Essentially, there are three steps: (1) extracting features in left and right images, respectively, using specific feature extraction operators; then (2) matching points of interest in the left and right images using a measure of similarity; for example, if corner information is extracted, then cross-correlation or LMS can be used as a measure of the similarity; and finally, (3) eliminating the outliers. By processing extracting features, the volume of data is reduced in order of magnitude compared to the original image data so that

one of the most important advantages of feature-based matching, as compared to area-based matching, is the reduction of the required computing time. Moreover, feature-based techniques are relatively insensitive to illumination changes between corresponding scenes - another advantage over area-based matching. As a result, it is very popular in computer vision and photogrammetry, and a lot of feature-based techniques can be found in the literature [Zhang et al., 1995; Liang and Heipke, 1996; Mount et al., 1999; Baillard and Dissard, 2000; Zhou and Shi, 2002].

2.4 Web-based Distributed Programming and Architectures

Distributed computing is one of the main ways of creating distributed application systems. It allows distributed processing of data and objects across a network of connected systems. This research uses distributed computing techniques to build a W3VPS.

A distributed computing system is a collection of a number of independent computers connected by a network. It provides the possibility of sharing information and peripheral resources by means of a communication network and established data-exchange protocols and processes. Distributed computing refers to computing activity which provides requested services to users by executing programs on a number of computers with data stored at one or more physical locations. The client/server model is the most widely used in constructing distributed applications and is a natural programming model for use in formalising the development of distributed computing.

2.4.1 Client/Server Model

As described in Sadoski [2000], a client/server model represents the relationship between two computer programs. In a client/server model, the client (one program) requests a service from the server (the other program), which fulfills the requested service and sends back a response which could be a processing result or a confirmation. In addition, all shared resources are held and managed by the server. Figure 2-8 illustrates a basic client/server model. This figure shows that, to receive a service, the user must send a service request to an individual server. Thus, in a network, the client/server model provides a convenient way to interconnect programs that are distributed across different locations. For a service to be accepted, both the client and server require a set of conventions that compose a protocol that must be implemented at both ends of a connection, and which must be understood by both, so that they are able to communicate with each other.

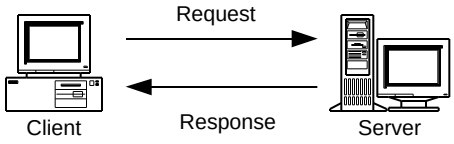

Figure 2-8. A basic client/server model

Traditionally, most client/server models use a 2-tier software architecture, which consists of three components distributed into two layers, client (requester of services) and server (provider of services). The three components are as follows [Sadoski, 2000]:

- User System Interface: accepting input such as text input and dialog and displaying the results.
- Processing Management: providing process services such as development, enactment, monitoring, and resource services.
- Database Management: providing data and file services such as accessing data stored on external storage devices.

The user system interface is allocated to the client tier; database management is placed on the server tier; and process management is placed on either client or server, or split between them. Figure 2-9 shows a typical two-tier client/server architecture.

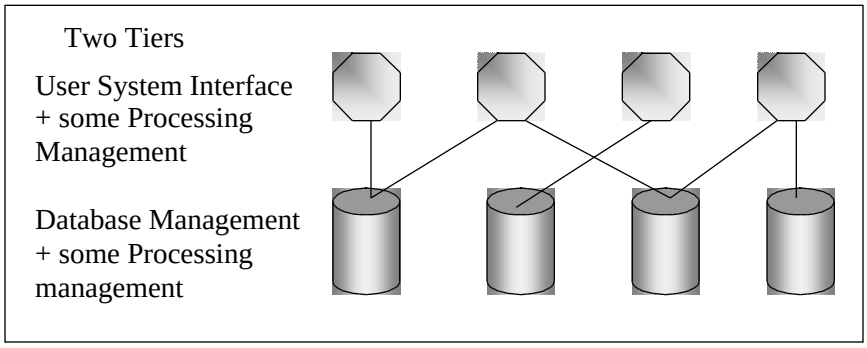

Figure 2-9. Architecture of two-tier client/server model (after Sadoski [2000])

Client/server two-tier software architectures are used extensively in non-complex systems in which the transaction load is light and processing rules (business rules) do not change very often. They require minimal operator intervention [Sadoski, 2000].

The World Wide Web is the most popular client/sever model. In a Web-based client/server model, the Web client can communicate with server software only through a Web server using Hypertext Transport Protocol (HTTP) [Sebesta, 2002]. A Web browser receives HTML/XML-based encoded information from the Web server. Data is stored on the Web server and is sent out to Web clients that request it. Web clients and servers have common features found in certain distributed computing environments: coarse-grained, distributed and heterogeneous, and message passing. Figure 2-10 illustrates a Web-based client/server architecture [Sebesta, 2002].

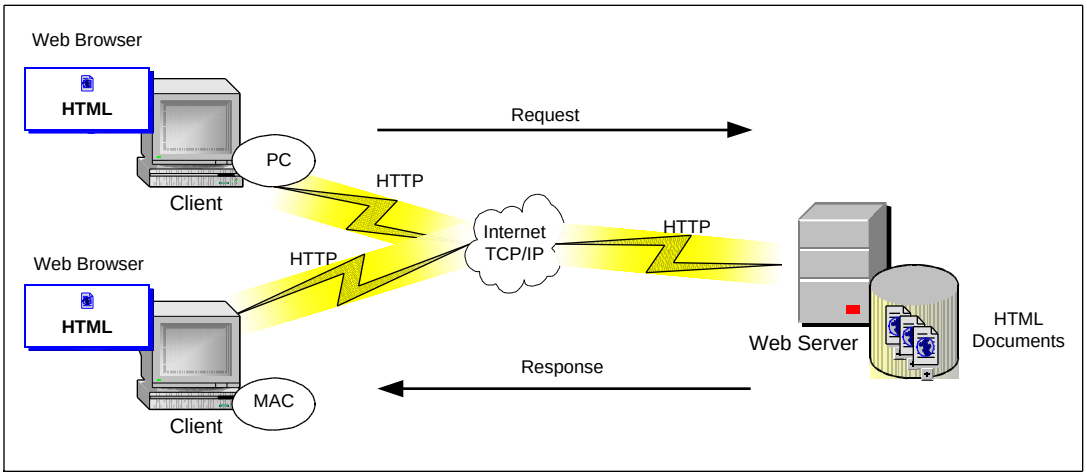

Figure 2-10. A Web-based client/server model (after Orfali et al. [1996])

2.5 Java Applet and Java Advanced Imaging (JAI)

Java Applet is Java technology for client side programming in client/server architecture of Internet computing. The JAI is Java technology for advanced image processing. In this research, Java Applet was used to build a rich GUI for 3D Web

visualization. To build a W3VPS, JAI was used to support general image processing such as displaying the image on the screen and obtaining the source image or images from image files or from another data source (e.g. a remote server). It was also incorporated into applets for a rich user interface of 3D display.

2.5.1 Java Applet

Applets are programs written in the Java language and are Java-based GUI components that are typically executed in a Web browser [Linden, 2002; Sun, 2002]. Applets can provide a powerful user interface for Web-based distributed applications [Linden, 2002; Sun, 2002]. They provide for client-side component portability in a heterogeneous Web environment and for the protection of the client machine.

2.5.2 Java Advanced Imaging

A simple rendering package suitable for rendering common HTML pages, but without the features necessary for complex imaging, can be found in an early version of the Java Advanced Windowing Toolkit (AWT) [Linden, 2002]. The Java 2D Application Program Interface (API) added support for more general graphics and rendering operations such as image rendering to extend the capabilities of the early AWT. However, a limited set of image processing operators is supported [Linden, 2002].

The JAI API further extended the Java platform by allowing sophisticated, high-performance image processing such as image tiling, region of interest, and

deferred execution to be incorporated into Java Applets and applications [Sun, 2003a]. Compared to other image solutions, JAI provides several advantages including cross-platform imaging, distributed imaging, object-oriented API, flexibility and extensibility, device independence, power, high performance, and interoperability (for the details, see [Sun, 2003a]).

Chapter 3

System Design and Development

A Web-based 3D Visualization Prototype System (W3VPS) for high-resolution satellite stereo images is developed in this research. The W3VPS allows the user to automatically generate 3D colour images without ground control points (GCPs), and to view the 3D effect of colour anaglyph images effectively through Web browsers without either special graphics hardware or high-speed Internet connections. The user can zoom in to view fine details of the high-resolution image in both 2D and 3D. This chapter describes the detailed design and development of the W3VPS.

3.1 System Overview

This system was developed based on the general concept of Zhang [2002a]. This general concept introduced in Chapter 1 formed the basis of the automatic W3VPS for colour 3D image generation and Web-based natural colour 3D visualization developed in this research. The W3VPS consists of three major components:

(1) Image-matching

- Automatic image-matching to evaluate correspondences for high-resolution pan-sharpened stereo image pairs

(2) 3D Image Generation

- Automatic epipolar image generation for effective 3D viewing
- Colour 3D image generation using anaglyph techniques

(3) 3D Web Visualization

- Automatic stereo matching between epipolar stereo image pairs
- Automatic image tiling for epipolar stereo image pairs
- Building GUIs as 3D viewers to display 3D images on local computers and through Web browsers, and to provide several services, for example, 3D effect refining
- Implementing the Web client/server architecture of 3D Web visualization to deploy the 3D image contents to the Web

Figure 3-1 illustrates the architecture of the W3VPS. The input to the W3VPS is a stereo pair of pan-sharpened high-resolution satellite images. The techniques for each component and data sources are described in the following sections.

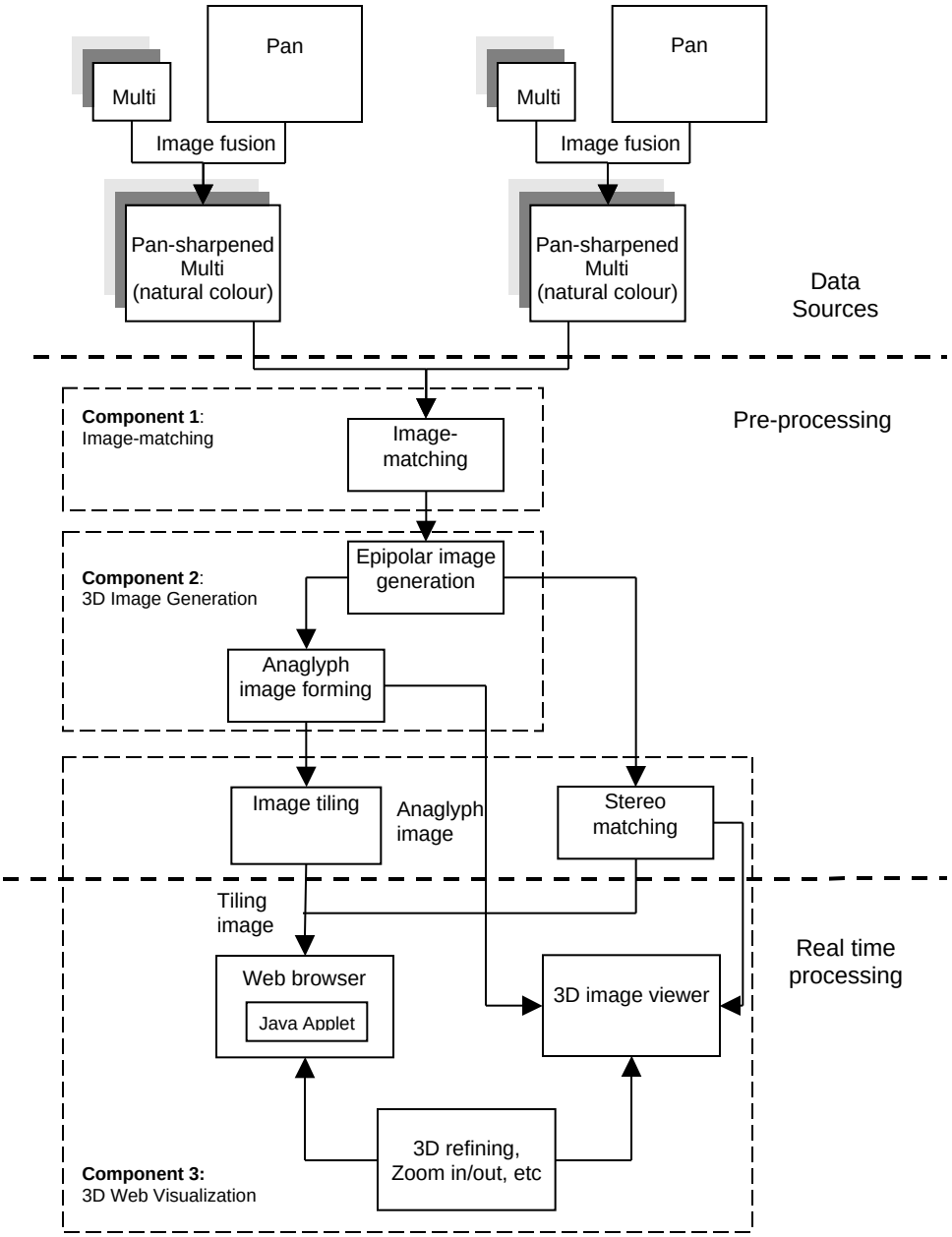

Figure 3-1. Architecture of the W3VPS
*Multi: Multi-spectral Image, *Pan: Panchromatic Image

3.2 Data Sources

The study area in this research is located in the city of Fredericton, Canada. IKONOS multi-spectral (MS) and panchromatic (Pan) stereo images taken in October 2001 are the source data. A stereo pair of pan-sharpened IKONOS natural colour images was used in this research; Pan-sharpened images were achieved by an automatic fusion algorithm, developed by Zhang [2002b] and integrated into PCI Geomatica.

A pan-sharpened image is a high-resolution colour image created by fusing together high-resolution Pan and low-resolution MS images. It preserves the original colour fidelity, thereby allowing for improved visualization and interpretation, as compared to a MS image. Figure 3-2 shows an example of the pan-sharpened MS image of the UNB area of Fredericton; Figure 3-3 shows an example of pan-sharpened high-resolution satellite colour stereo images of downtown Fredericton.

1m Pan image + 4m MS image →1m pan-sharpened image

Figure 3-2. An example of pan-sharpened MS image generation
(Location: UNB, Fredericton)

Figure 3-3. An example of pan-sharpened high-resolution MS satellite stereo images (Location: downtown Fredericton)

3.3 Image-matching

A lot of work has been done on automatic image-matching in aerial and satellite stereo image applications such as DEM generation. Feature-based and area-based matching are two very common techniques. The automatic image-matching method developed in this thesis combines advantages realized by both feature-based and area-based techniques described in Chapter 2. This new automatic method also uses the epipolar constraint to eliminate outliers, and incorporates the epipolar image re-sampling method. An overview of the algorithm is illustrated in Figure 3-4.

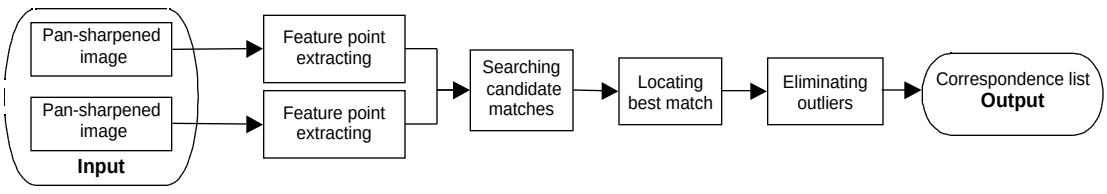

Figure 3-4. Overview of the image-matching process

Certain parameters, such as the matching window size and the threshold value, must be pre-determined. Because factors, such as the resolution and quality of the image, influence the chosen parameters, it is non-trivial to decide upon appropriate parameters. The satellite stereo images processed are too large (more than 100Mb for an image of size 4000 by 3000 pixels in .tif format). It is unproductive and inefficient for users to attempt to alter the parameters to overcome the effects of large image size. In this research, two different components have been developed to address this problem and to enable users to efficiently determine the optimal parameters for their particular stereo pair in an interactive process. Component 1 includes the first three steps of the automatic image-matching processing illustrated in Figure 3-4 while component 2 includes all steps.

Component 1 is called the test component: as discussed in section 2.3.2, there is no accepted standard for the optimal template size of the correlation, so trying different template size can be a compromise to solve the problem. Thus, the component's purpose is to allow users using small parts of stereo images as a test stereo pair to determine optimization parameters for further processing of the entire image. GUIs are provided so that users can interactively experience the result of each processing step of image-matching directly, and then make a reasonable selection of matching parameters (Figures 3-5, 3-6, 3-7, and 3-8).

Figure 3-5. A GUI for Sigma selection and displaying corner images

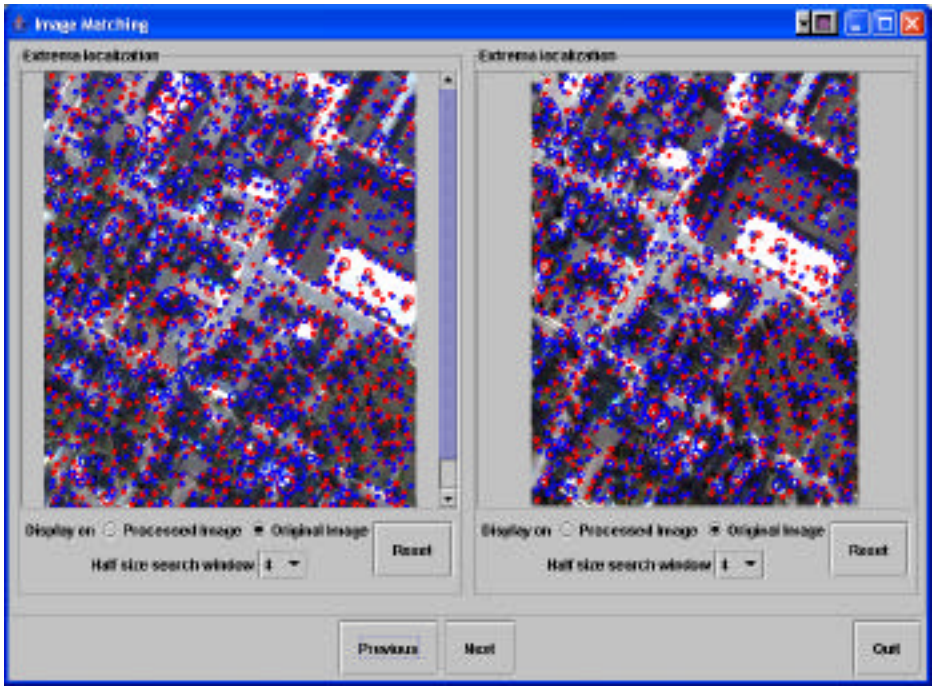

Figure 3-6. A GUI for search window size selection and displaying extrema points

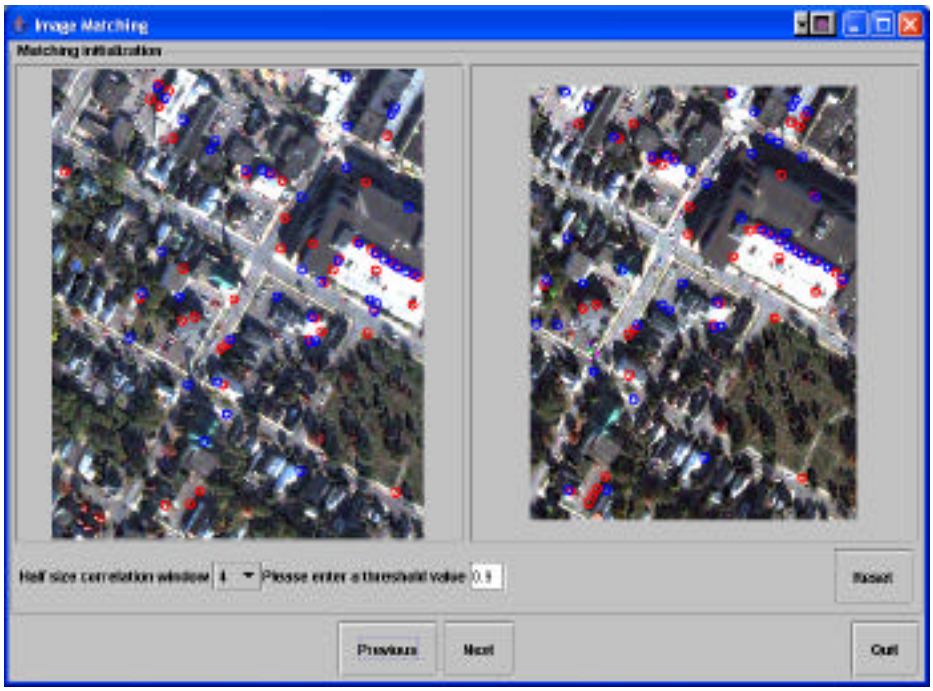

Figure 3-7. A GUI for the search window size and threshold value selections of the initial matching and displaying candidate match points

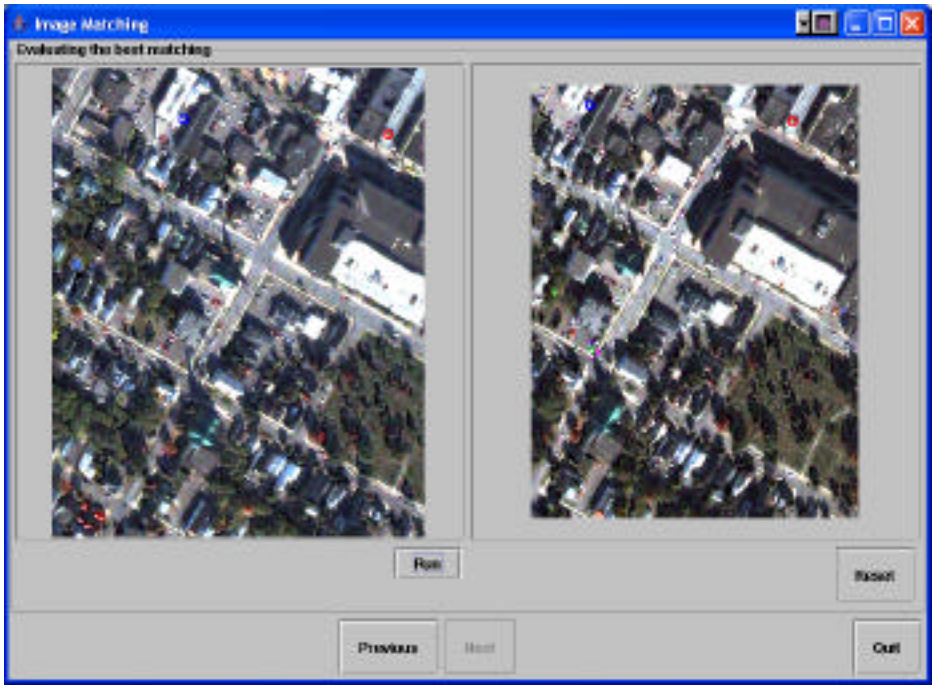

Figure 3-8. A GUI for displaying the result of locating the best match

Component 2 is called the batch component: users can utilize the optimization parameters determined in Component 1 to automatically process an entire large stereo image pair. A GUI is provided to allow users to enter parameters for the matching process (see Figures 3-9).

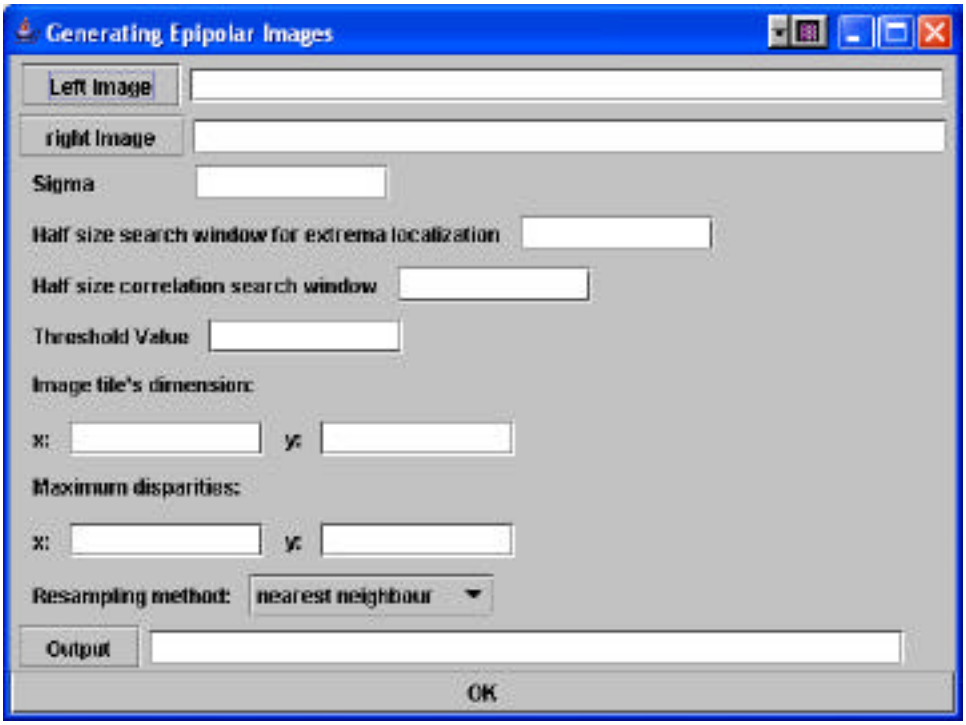

Figure 3-9. A GUI for entering the parameters of the image-matching process

3.3.1 Image Tiling

Previous image-matching algorithms [Zhang et al., 1995; Baillard and Dissard, 2000; Zhou and Shi, 2002] focused on the quality of the match result and incorporated automatic operation. To date, few algorithms are concerned with algorithm efficiency or memory requirements – what would seem to be important

criteria, given the large data volume present in high-resolution satellite colour stereo images. Therefore, in this research, the requirements of memory and algorithm efficiency have been adopted as important design considerations in producing a suitable and efficient image-matching algorithm for the W3VPS. The even distribution of correspondences over the whole image area as an important factor for establishing accurate epipolar geometry is another matching algorithm design consideration. In this research, image-tiling techniques are applied to meet these factors.

Handling large digital images in their entirety is rather cumbersome. The term “tiling” refers to partitioning the original image into rectangular blocks (tiles), which are processed independently, as though they were entirely distinct images. It transforms the matching problem into a task of dealing with image tiles instead of a whole image to improve system efficiency. The W3VPS deals with tiled stereo images in successive stages of the image-matching process, which means that the operations - including feature point extraction, candidate match searching, and best match locating - are performed independently on each different tiled stereo image pair (see Figure 3-10). A tiling process creates these tiled stereo image pairs. This process partitions the original left and right images into rectangular blocks having the same dimensions, apart from those on the boundaries of the image. Each tile in the left image corresponds to a distinct tile in the right image, forming a tiled stereo image pair. However, the sizes of the right image tiles are larger than those of the left image tiles. There are overlaps between the neighbouring tiles of the right image

but not those of the left image. This ensures that the area of overlap between the tiled stereo image pairs compensates for disparities between images. As expected, tiling affects the result and efficiency of stereo matching. Having overly small tiles creates some tiled stereo image pairs without overlapping areas. Conversely overly large tiles cause less efficient performance of stereo matching. In the W3VPS, a GUI is provided for users to select the tile's dimension (see Figure 3-9); the default dimension is 250 by 250 pixels.

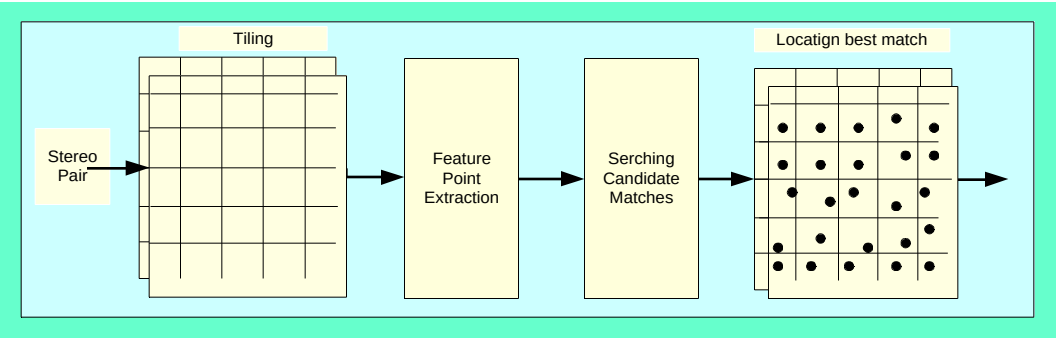

Figure 3-10. Tiling, feature point extraction, searching candidate matches, and locating the best match

The main advantages of the tiling process include: the reduction of memory requirements and the easy implementation of the match algorithm; it can also prevent the established correspondences from being very close to each other for epipolar re-sampling because, for each tiled stereo image pair, only the best match is chosen. In the implementation phase, this algorithm can be speeded up considerably by means of parallel computing because the processing for each tiled stereo image pair can be done independently.

3.3.2 Feature Point Extraction

The detection of feature points is an essential task for many applications such as 3D scene reconstruction and DEM generation from stereo image pairs. The corner as an intuitive feature point shows a great change in intensity in 2D. Therefore, corners are naturally distinguished from the neighbouring points. Because of their existing high location information content, corners can be reliably matched between images; therefore corner detectors have been broadly used as feature point detectors.

One of the early corner detection approaches involved first extracting the edges as a chain code, then identifying the corners as points which have maximal curvatures [Asada and Brady, 1986; Medioni and Yasumuto, 1986]; or first segmenting the image into regions, extracting the boundary as chain code, then searching for the line segment intersections [Horaud et al., 1990]. However, this type of approach is seldom used because it depends on the previous segmentation or edge detection step and is also time-consuming. Subsequently developed corner detectors, reported in the literature in the last few years, may be divided into two groups: (1) curvature based; and (2) interest operators or feature point detectors.

A corner detector based on the curvature of the grey-level image is applied in the matching process to extract high curvature points for incorporation due to the relatively low computational cost and strong citation record in some applications.

3.3.2.1 Algorithm Description

The algorithm used for extracting points of interest in the image-matching process typically involves two steps:

(1) For each stereo image tile, calculate the corner image *g(x, y)* for highlighting corners; and then

(2) Detect the corner points in each corner image.

The output grey level *g(x,y)* of the corner image is defined as the product of the isophote curvature and the absolute gradient magnitudes as follows:

$$g(x,y) = k(x,y) \quad | \quad I(x,y)|^3 \qquad 3\text{-}1$$

$$k(x,y) \approx \frac{\partial^2 I}{\partial t^2} / | \quad I(x,y)|, where \quad I^2 >> 1 \qquad 3\text{-}2$$

where *I(x, y)* denotes the original image; κ*(x, y)* is the curvature of the iso-intensity line passing through the point; and *I(x,y)* is the local gradient vector. Corners are highlighted when the output grey level is high, which occurs in cases where both the local gradient and the curvature are high yielding. The advantage of using this measure is that the product will focus on high isophote curvatures close to high contrast edges.

Let *n* be the edge normal and *t* be the edge tangential. Differentiating *I(x, y)* in directions of *n* and *t* yields:

$$\frac{\delta^2 I}{\delta n^2} = \frac{1}{| \quad I(x,y)|^2}\left(I_x^2 I_{xx} + 2I_x I_y I_{xy} + I_y^2 I_{yy}\right) \qquad 3\text{-}3$$

$$\frac{\delta^2 I}{\delta t^2} = \frac{1}{| \quad I(x,y)|^2}\left(I_x^2 I_{yy} - 2I_x I_y I_{xy} + I_y^2 I_{xx}\right) \qquad 3\text{-}4$$

where I_x and I_y indicate the *x* and *y* directional derivatives respectively. The Gaussian function is a well-known operator in image processing for reducing high frequency image information to smooth the image. The most popular masks used for

computer vision are those that combine simple directional operators with Gaussian smoothing. In the implementation of this research, derivatives of a Gaussian distribution are used as the convolution filters that comprise the operator. Gaussian standard deviation (σ) is used as the parameter to adjust the size of the highlighted features. Letting *G* be the Gaussian function, a 2-dimensional Gaussian function can be expressed as follows:

$$G_\sigma = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}5$$

A discrete smoothing kernel is obtained by constructing a *(2k+1)*(2k+1)* array whose *i, j*th value is

$$H_{ij} = \frac{1}{2\pi\sigma^2} e^{-\left((i-k-1)^2+(j-k-1)^2\right)/2\sigma^2} \qquad 3\text{-}6$$

Derivatives of the Gaussian function applied to the image tile are calculated as follows:

$$I_x = \frac{\partial G_\sigma(x,y)}{\partial x} = \frac{x}{\sigma^2}\frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}7$$

$$I_y = \frac{\partial G_\sigma(x,y)}{\partial y} = \frac{y}{\sigma^2}\frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}8$$

$$I_{xy} = \frac{\partial G_\sigma(x,y)}{\partial xy} = \frac{xy}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}9$$

$$I_{xx} = \frac{\partial^2 G_\sigma(x,y)}{\partial x^2} = \frac{x^2-\sigma^2}{\sigma^2}\frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}10$$

$$I_{xy} = \frac{\partial^2 G_\sigma(x,y)}{\partial y^2} = \frac{y^2-\sigma^2}{\sigma^2}\frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \qquad 3\text{-}11$$

Analogous discrete smoothing Gaussian kernels can be given for I_x, I_{xx}, I_{xy}, I_y, and I_{yy}. Once the derivatives are known, the values of the corner image are easily computed. Because the Gaussian function (and its derivatives) can be separable,

using their separability property becomes one of their intrinsic practical advantages in computer vision or photogrammetry [Boomgaard et al., 2001]. In practice, it exploits the characteristic that the convolution of an image *I(x, y)* with a separable kernel is equivalent to the composition of *d* convolutions each using a one-dimensional convolution kernel [Boomgaard et al., 2001], thus speeding up computation. In this research, for a more computationally efficient solution, each discrete smoothing kernel of Gaussian derivatives is separable into the outer product of two one-dimensional vectors during the implementation phase.

Care must be taken with σ (Sigma); the units of Gaussian standard deviation (σ) are interpixel spaces, referred to as pixels here. The qualitative analysis of σ yields the following [Forsyth and Ponce, 2003]: (1) If σ is too small, the smoothing will have little effect; (2) For a large σ, this will give a good estimate of the pixel's value; noise will largely disappear at the risk of some blurring; image detail will disappear; and (3) if σ is large, the kernel size must be large also. For this reason, a 2-dimensioanl Gaussian function (and its derivatives) was applied using a (3σ +1)(3σ +1) kernel window, as advised by Forsyth and Ponce [2003]. The defaulted Gaussian standard deviation is defined as 1 in the W3VPS. Users can also define it through a GUI (see Figure 3-9). The test component of image-matching provides a GUI to allow users to interactively experience the result of corner images with different σ and then to decide on an optimal σ for their particular images (see Figure 3-5). Figure 3-11 shows an original image and its corner image that highlights corners in white or black.

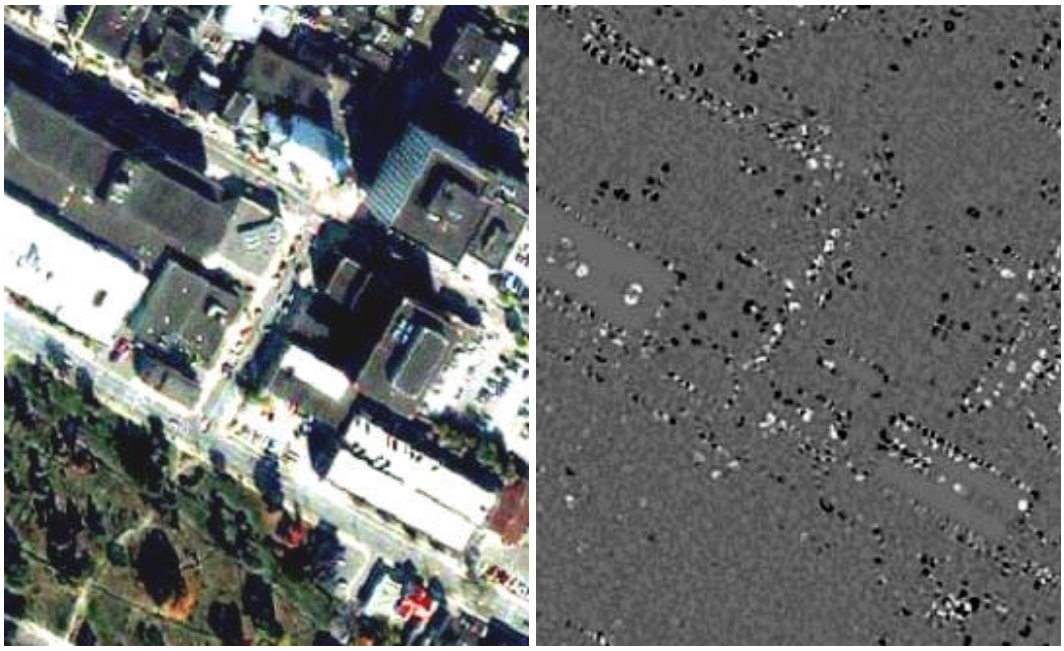

Figure 3-11. Results of the corner image generation
Left: original image. Right: corner image

After corner images are obtained (that is, the highlighting of corners), the system performs the local extrema detection which detects positive as well as negative extrema of the corner image. It is worth noting that the corner is located where both the local gradient and the curvature are high. Geometrically, corners correspond to the annihilation or creation of a saddle with either a minimum or a maximum. To extract the corner, an elliptical area (search window) of size *n* by *m* (in this research, we define a size of *2n+1* by *2n+1*) is defined. Marking points where the area of this ellipse is extremal could mark the corner. Local extrema detection is as follows: (1) centering the elliptical area at a point; (2) looking for image neighbourhoods of this point; if all of the neighbourhood's magnitudes are

lesser or greater than that of the point, this point is marked as a feature point (a corner).

The test component of image-matching provides a GUI to allow users to experience the results of corner detection with different window sizes and then to select an optimal size (Figure 3-6). Corner detection is based on the threshold of the absolute value of extrema of this operator. Figure 3-12 shows an original image and another with its image with extracted corner information.

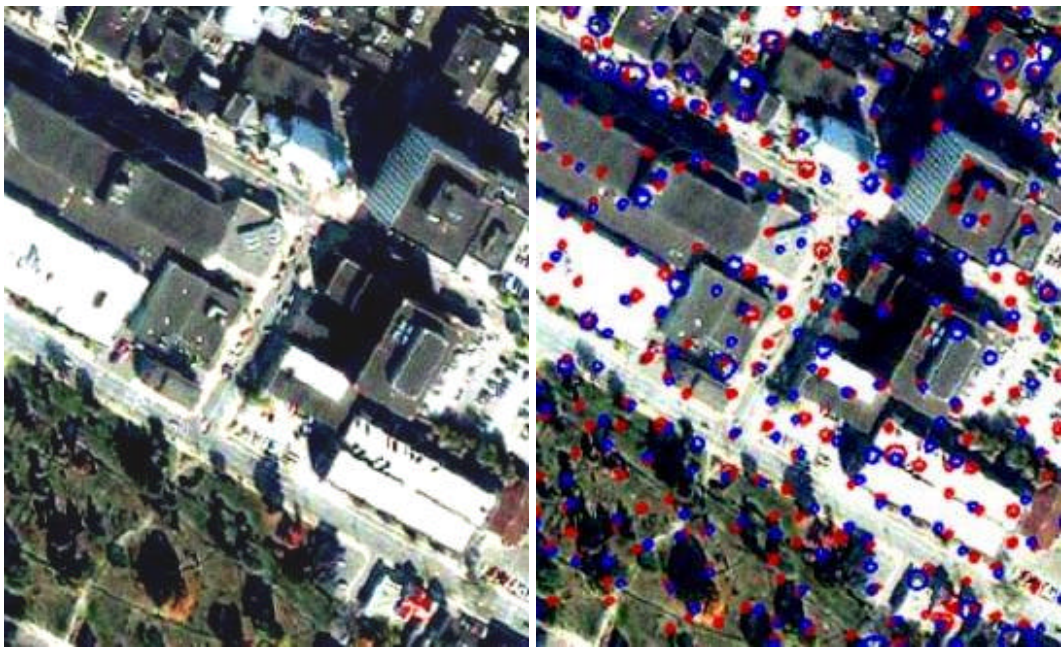

Figure 3-12. Results of the corner extraction process. Blue or red circles correspond to maxima or minima, respectively. The circles sizes reflect the maxima or minima magnitudes.

3.3.3 Searching for Candidate Matches

Searching for candidate matches finds possible matches in the right image tile for each high curvature point of the left image tile. After extracting points of

interest, the system performs the initialization of the matching through a correlation technique to decide whether two feature points are a candidate match. Each point of the source (left) image tile is associated with possible matches in the target (right) image tile. For each given point p_1 (source point) in the left tile, we define a correlation window (n by m, image patch) centered at this point. Then, we perform correlation operations between p_1 and all extracting points (target points) of interest. At this time, maxima or minima of the source image can be matched only with maxima or minima of the target image. The normalized cross-correlation coefficient between the source point and target point is calculated by Equation 2-12.

The cross-correlation coefficient is used as a constraint to select the most consistent matches. A given threshold input through a GUI (see Figure 3-9) determines candidate matches. If the cross-correlation value of a given couple of points is higher than the threshold, it is considered a candidate match.

Before calculating the cross-correlation value of the two points, the disparity constraint, representing *a priori* knowledge about the disparities between the matched points, is used to reduce the search area from a whole image to a search window (Figure 3-13). Thus, two points comprise a possible match if they are neighbours within a disc of radius R. There are two ways to obtain the disparity constraint information: (1) disparity constraint defined by users such as the minimum and maximum disparity information; and (2) disparity information obtained from the neighbouring tiled stereo image pairs. The idea of the latter is to propagate the continuity constraint through the neighbourhood (image tiles).

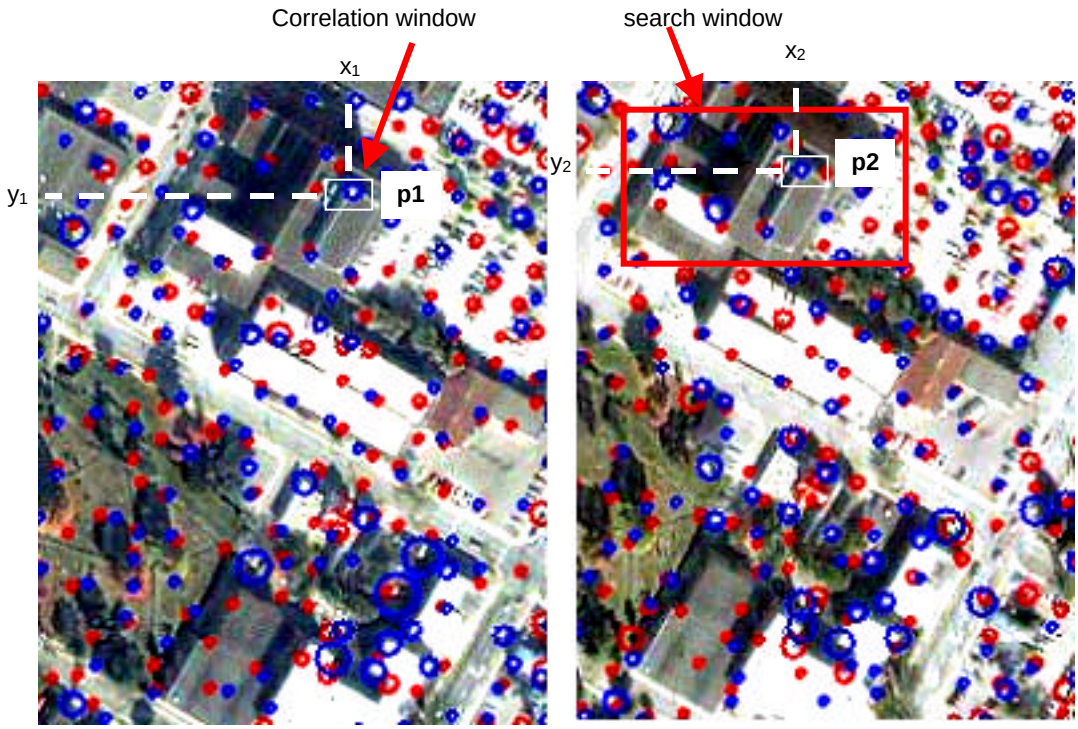

Figure 3-13. Correlation: constraint search window for left and right images. The kernels centred at point $p_1(x_1, y_1)$ in the left image and $p_2(x_2, y_2)$ in the right image

As described in Chapter 2, the optimal size of the matching window is important. Unfortunately, there is no accepted standard for the optimal size of a matching window, and so a compromise such as trying different template sizes in turn must be found. In this system, the test component of image-matching provides a GUI to allow users to experience the results of candidate matching with different window sizes and threshold values (Figure 3-7). At this time, there still are some ambiguous matches among the candidate matches, especially when there is a high density of feature points. Figure 3-14 shows the result of searching for candidate matches in downtown Fredericton.

Figure 3-14. Results of searching for candidate matches
Blue or red circle pairs correspond to maxima or minima matches, respectively.

3.3.4 Evaluating the Best Match

A few established correspondences between images are sufficient for epipolar image re-sampling.To this end, during the process of evaluating the best match, the best match among the set of candidate matches for each tiled stereo image pair is selected. This process is based on the measure of candidate matches lying in the neighbourhood of each point being analyzed [Zhang et al., 1995]. Zhang et al. [1995] defined the strength of match (SM) as a means of evaluating candidate matches. Zhang's SM method [Zhang et al., 1995] considers close candidate matches to be superior to distant ones. However, when most candidate matches of tiled stereo image pairs join closely together, it may result in false matches (see Figure 3-15 Middle)

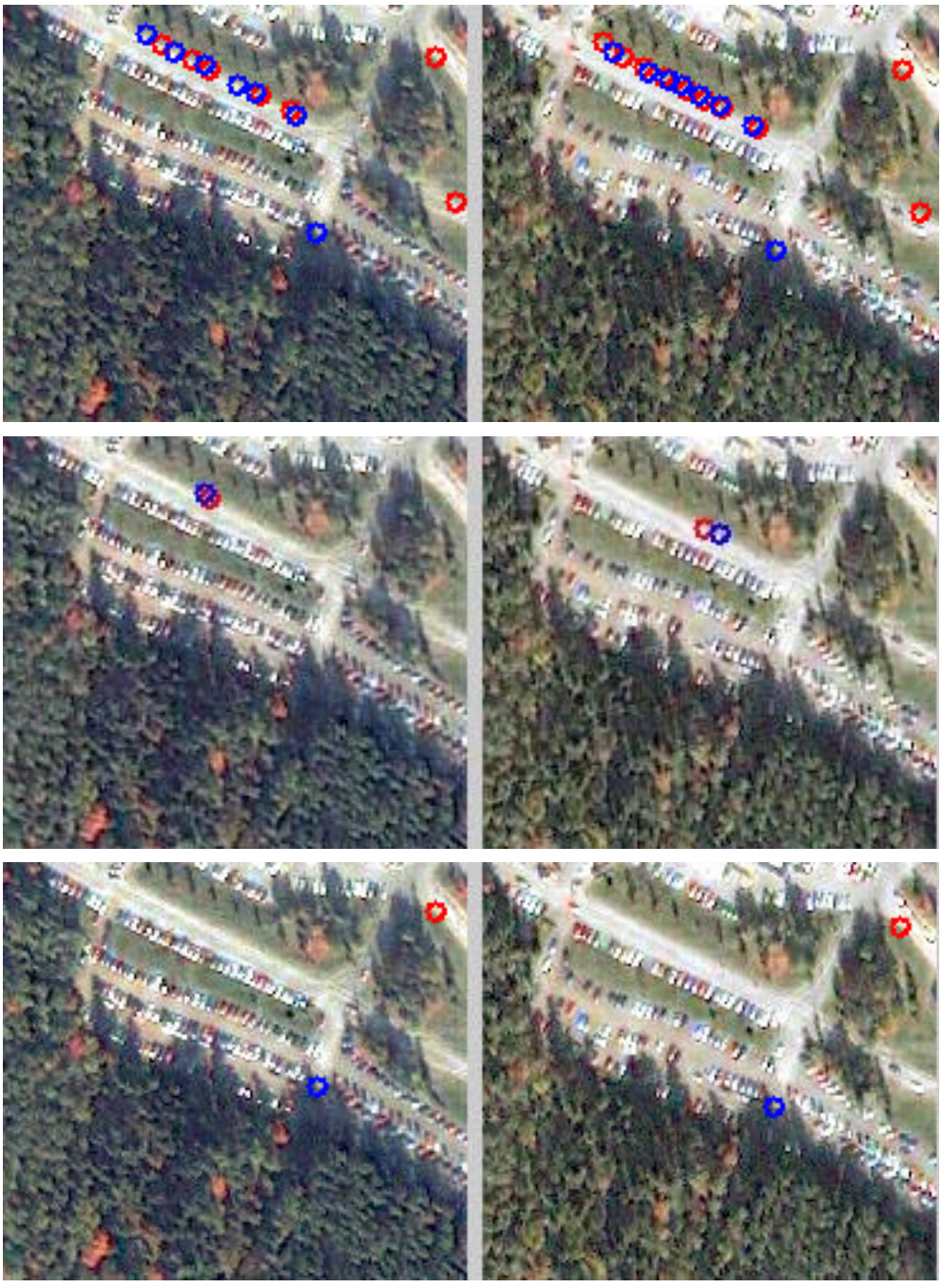

Figure 3-15. Automatically locating the best match in a tiled stereo image pair.
Top: Candidate match information;
Middle: Best match information measured by Zhang's SM;
Bottom: Best match information measured by the modified SM.

In this research, therefore, a modified measure SM is developed to better locate the best match. The idea of the modified SM is to allow candidate matches to measure themselves by using the continuity constraint throughout the neighbourhood. This is similar to that of Zhang's SM; however, the modified SM assumes that all candidate matches have the same weight. Figure 3-15 (bottom) shows the result of the modified SM method. In this research, the proposed modified SM of points of candidate matches is defined by the following expression which is a slight change from Zhang's SM:

$$S_M(m_{1i}, m_{2j}) = c_{ij} \times \sum_{n_{1k} \in N(m_{2i})} \left[\max_{n_{2l} \in N(m_{2j})} c_{kl}\, \delta(m_{1i}, m_{2j}; n_{1k}, n_{2l}) \right] \qquad 3\text{-}12$$

where m_{1i} and m_{2j} are feature points of the source and target image tiles of candidate matches; $N(m_{1i})$ and $N(m_{2j})$ are the neighbourhood areas of m_{1i} and m_{2j} with a disc of radius R, respectively; and n_{1k} and n_{2l} are the feature points of the source and target image tiles of candidate matches lying over $N(m_{1i})$ and $N(m_{2j})$, respectively. Coefficients C_{ij} and C_{kl} can be cross-correlation coefficients of candidate matches (m_{1i}, m_{2j}) and (n_{1k}, n_{2l}), which characterize similar levels of the candidate matches. $dist(m_{1i}, m_{2j}; n_{1k}, n_{2l})$ is average distance between two points of two candidate matches in source and target image tiles where $dist(m, n)$ is the Euclidean distance between points m and n. In this system, the set of image tiles are defined as the neighbourhood areas in this research because, for each tiled stereo image pair, only one match is selected; accordingly, C_{ij} and C_{kl} are defined as 1.

$$dist(m_{1i}, m_{2j}; n_{1k}, n_{2l}) = [d(m_{1i}, n_{1k}) + d(m_{2j}, n_{2l})]/2 \qquad 3\text{-}13$$

$$\delta(m_{1i}, m_{2j}; n_{1k}, n_{2l}) = \begin{cases} \exp(-r/\varepsilon), & if \quad r < \varepsilon \\ 0, & otherwise \end{cases} \qquad 3\text{-}14$$

where r is the relative distance difference; and ε is the threshold on r. In this system, ε is defined as 0.3, the same value used by Zhang et al. [1995].

$$r = \frac{\left| d(m_{1i}, n_{1k}) - d(m_{2j}, n_{2l}) \right|}{dist(m_{1i}, m_{2j}; n_{1k}, n_{2l})} \qquad 3\text{-}15$$

The Match strength algorithm is analogous to having a poll where the candidate match who garners the most supporters wins. Thus, in this research, the strategy of choosing the best match is the "winner-takes-all-matches" which means that any match with the highest SM is selected to be the best match. Figure 3-16 shows the results of evaluating the best match process. After all of the matches have been established for all tiled stereo image pairs, a correspondence list is obtained for the entire stereo image.

Figure 3-16. Results of evaluating the best match.
Blue or red circle pairs correspond to maxima or minima matches, respectively.

3.3.5 Eliminating Outliers

The outliers will affect the precision of epipolar image re-sampling if we directly use the results from the previous image-matching processes. This procedure addresses the outlier problem and processes the whole image; in fact, it deals with a set of correspondences. In all of the matches established so far, as described in section 3.3.1 through section 3.3.4, there still exist outliers even when a SM threshold is used. This is due to: (1) The localization error of corners; and (2) False matches. False matches exist because, for most cases, there exist some homogenous areas with simple features (such as water area bridge features) (see Figure 3-17); additional false matches occur because the stereo images were created at different times (several days difference). In this research, the proposed solution for this problem is an evaluation of the vertical disparity.

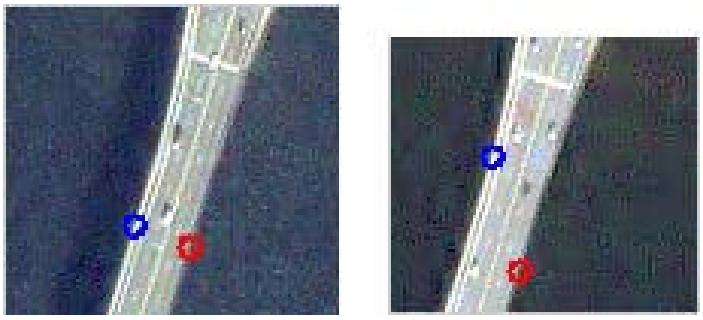

Figure 3-17. Examples of false matches

Currently, computer vision researchers have proposed using robust regression techniques (Least Median of Squares (LMedS)) combined with recovery of the epipolar geometry to eliminate the outliers [Shapiro and Brady, 1995; Zhang

et al., 1995; Chua et al., 2000; Zhou 2002]. This approach is reliable and appropriate when the outlier subset is large or when outliers control structure. However, the LMedS problem cannot be simplified to a weighted least-squares problem and, so, it is complicated to implement, and a robust regression approach entails a significant computational overhead. In this thesis, a new approach is developed to reject false matches through a regression diagnostics approach achieved through an iterative process. This process, for each loop, computes an estimate of the epipolar re-sampling over all matches, and then checks vertical disparities of all matches between two epipolar re-sampling images. The epipolar re-sampling method used in this research and the new algorithm for eliminating outliers are described below.

3.3.5.1 Epipolar Re-sampling of High-resolution Linear Pushbroom Satellite Imagery

The geometric characteristics of satellite linear pushbroom images are different from those of perspective images [Otto, 1988; Kim, 2000], and this makes conventional methods of epipolar re-sampling unsuitable for them. Otto [1988] reported that strict epipolar images could not be generated from SPOT imagery without DTMs. This is also true for high-resolution satellite imagery produced by the CCD line scanner. Because of this, several methods were developed to generate pseudo epipolar images by using DTMs [O'Neil and Dowman, 1988; Otto, 1988; Haala et al., 1998]. Even though DTMs already exist from an ultimate source, those methods are complicated to implement and cannot reach a practical level of accuracy without highly precise orientation parameters. Ono [1999] proposed a

method for the epipolar re-sampling of satellite imagery using a 2D affine orientation model without the rigorous orientation parameters or DTMs. This method is based on the assumptions that the satellite altitude parameters are almost constant during the collection of high-resolution imagery and that its movement is linear. The author reported that this method is very simple and highly accurate and that it is appropriate for real time mapping of satellite images. In this research, the 2D affine orientation model is, therefore, used to generate epipolar images.

3.3.5.1.1 2D Affine Orientation Model

Okamoto et al. [1998] had proposed an orientation model of CCD line sensor imagery based on an affine projection. This orientation model, called the 2D affine orientation model, is derived from conventional collinearity equations (Equations 2-4 and 2-5) under the hypothesis that the central projection can be approximated by the parallel projection for high-resolution satellite imagery with a small field angle. The collinearity equation for the 3D analysis of affine line scanner imagery can be given as:

$$\begin{pmatrix} 0 \\ y \\ -c \end{pmatrix} = \lambda \left(R_{\phi i} R_{\omega i} R_{\kappa i} \right)^T \begin{pmatrix} X - X_{oi} \\ Y - Y_{oi} \\ Z - Z_{oi} \end{pmatrix} \qquad 3\text{-}16$$

where i is the line number; and (X, Y, Z) are ground coordinates of an object point. Coordinates of project centre X_{oi}, Y_{oi}, Z_{oi} and angles ϕ_i, ω_i, κ_i express exterior orientation parameters; $R_{\phi i}$, $R_{\omega i}$, $R_{\kappa i}$ are rotation matrixes and λ is a scale parameter. c expresses the principal distance; and y is the image coordinate of an object point.

Equations of the affine transformation collinearity relationship from 3D to 2D can be expressed as follows:

$$x = D_1X + D_2Y + D_3Z + D_4 \qquad 3\text{-}17$$
$$y = D_5X + D_6Y + D_7Z + D_8 \qquad 3\text{-}18$$

where *Di* (i=1, …, 8) are independent coefficients; *(x, y)* are 2D affine image coordinates of an object point; and *(X, Y, Z)* are ground coordinates of an object point.

Because the actual satellite images are taken from a perspective along the scanning line, they are not, strictly speaking, affine images. For rigorous analysis, images should be transformed into affine images. The goal of this system is to automatically generate and to display 3D images on screens or printouts, so that the transformation from original images to affine images is ignored for the sake of simplifying the problem and for computational efficiency. In this research, the epipolar geometry of affine imagery is directly used on satellite images to generate the epipolar images under the hypothesis described at the start of this section.

3.3.5.1.2 Epipolar Re-sampling for Affine Images

As shown in Figure 3-18, affine projection, in which the viewing rays are parallel to each other, is a special case of the centre perspective projection with a focal length of infinity. Thus, the epipolar geometry of an affine image can be defined as that of a perspective image. The epipolar lines and epipolar planes are defined as shown in Figure 3-19. Figure 3-20 shows the approach used to generate

the epipolar images. It can be achieved by parallel projecting the affine images into the same virtual plane.

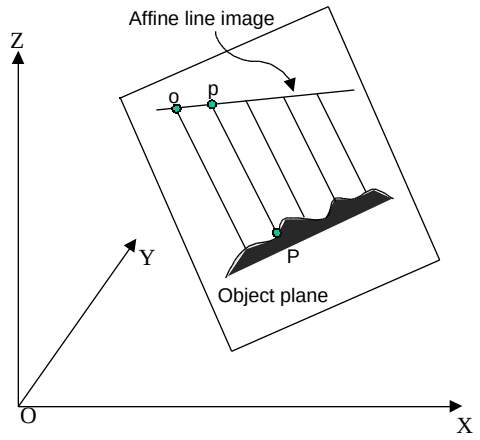

Figure 3-18 Image geometry of parallel ray projection with respect to O-XYZ (after Okamoto et al. [1998]).
*O-XYZ: object space coordinate system

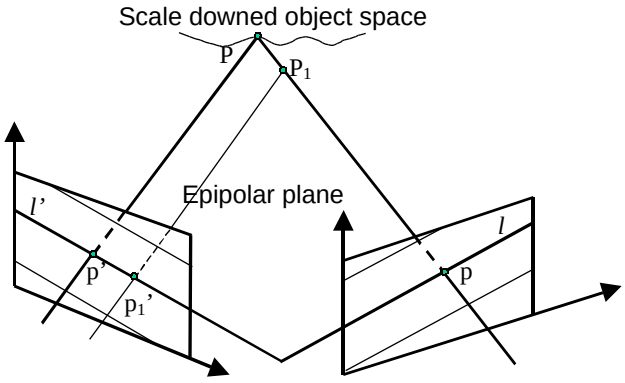

Figure 3-19. Affine epipolar geometry: For two given parallel projection images, the epipolar plane is defined by a point p (right image) and two projection directions (Pp' and P_1p_1' in left image). The intersections between the epipolar plane and right or left images are the epipolar lines *l* and *l'*, respectively (after Ono [1999]; Forsyth and Ponce [2003]).

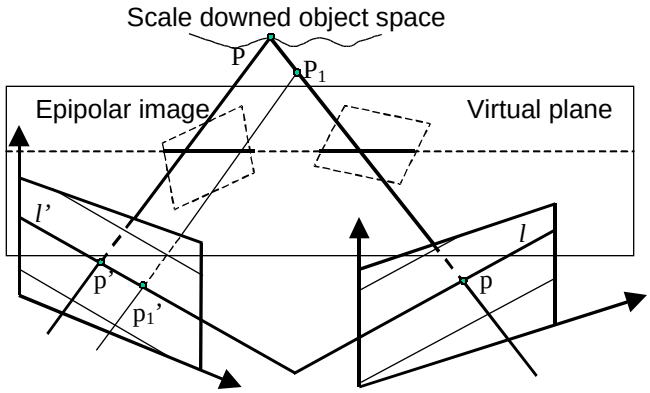

Figure 3-20. Affine image epipolar re-sampling (after Ono [1999])

In the implementation phase, let the left image be the virtual plane. Thus, it requires only the projection of the right image onto the virtual plane. It is the same as the affine transformation of the right image to the left image. The relationships between left image coordinates (x_l, y_l) and right image coordinates (x_r, y_r) are given by Equations 3-19 and 3-20.

$$x_r = M_1 x_l + M_2 y_l + M_3 \qquad 3\text{-}19$$
$$y_r = M_4 x_l + M_5 y_l + M_6 \qquad 3\text{-}20$$

where M_i (i = 1,…,6) are independent coefficients and can be solved through more than three known corresponding points. After parallel projections, the epipolar line direction should be determined by algebraic solutions. According to the basic equations of affine images (Equations 3-17 and 3-18), equations of the affine transformation collinearity relationship of affine stereo images can be expressed as follows:

$$x_l = K_{1l} X + K_{2l} Y + K_{3l} Z + K_{4l} \qquad 3\text{-}21$$
$$y_l = K_{5l} X + K_{6l} Y + K_{7l} Z + K_{8l} \qquad 3\text{-}22$$

$$x_r = K_{1r}X + K_{2r}Y + K_{3r}Z + K_{4r} \qquad 3\text{-}23$$
$$y_r = K_{5r}X + K_{6r}Y + K_{7r}Z + K_{8r} \qquad 3\text{-}24$$

By eliminating the X, Y, and Z from Equations 3-21 through 3-24, a single equation expressing the epipolar line of affine images in terms of image measurements is described as follows:

$$y_l = A_1 x_l + A_2 y_r + A_3 x_r + A_4 \qquad 3\text{-}25$$

where A_i (i = 1,…,4) are coefficients. To solve Equation 3-25, more than four corresponding points are required. The epipolar line direction (Rotation angle φ) corresponding to A_1 is arctan(A_1). The least-square approximation technique, also known as regression, is used to calculate coefficients A_i of Equation 3-25 in this system. The least-squares solution is given by:

$$\hat{\lambda} = \left(\Phi P \Phi^T\right)^{-1} \Phi P l \qquad 3\text{-}26$$

where $\hat{\lambda}$ is the coefficient vector (solution vector); Φ is a Vandermonde's matrix (Equation 3-27 is the transposed matrix of the Vandermonde's matrix); P is a weighed matrix to be an identity matrix I in this system; l is an observation vector; and (x_{li}, y_{li}) and (x_{ri}, y_{ri}) are coordinates of the left and right images of correspondences, respectively.

$$\Phi^T = \begin{bmatrix} x_{l1} & y_{r1} & x_{r1} & 1 \\ x_{l2} & y_{r2} & x_{r2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{\ln} & y_{rn} & x_{rn} & 1 \end{bmatrix} \qquad 3\text{-}27$$

$$\hat{\lambda} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} \qquad 3\text{-}28 \qquad\qquad l = \begin{bmatrix} y_{l1} \\ y_{l2} \\ \vdots \\ y_{ln} \end{bmatrix} \qquad 3\text{-}29$$

The general procedures of epipolar image re-sampling are summarized as follows:

1. Calculate the affine transformation coefficients (Equations 3-19 and 3-20)
2. Carry out the affine transformation
3. Calculate the rotation angle ϕ
4. Rotate the stereo images by an angle ϕ

As described so far, to generate epipolar images, more than four correspondence coordinates are sufficient without any other information.

3.3.5.2 Algorithm for Eliminating Outliers

The basic idea of this algorithm comes from the properties of the epipolar geometry of linear pushbroom images and combines with the epipolar image re-sampling method. The algorithm presented in this thesis is shown in Figure 3-21 in a general framework of the epipolar line estimated by regression, and uses a new regression diagnostics based on the vertical disparity between an epipolar stereo image pair.

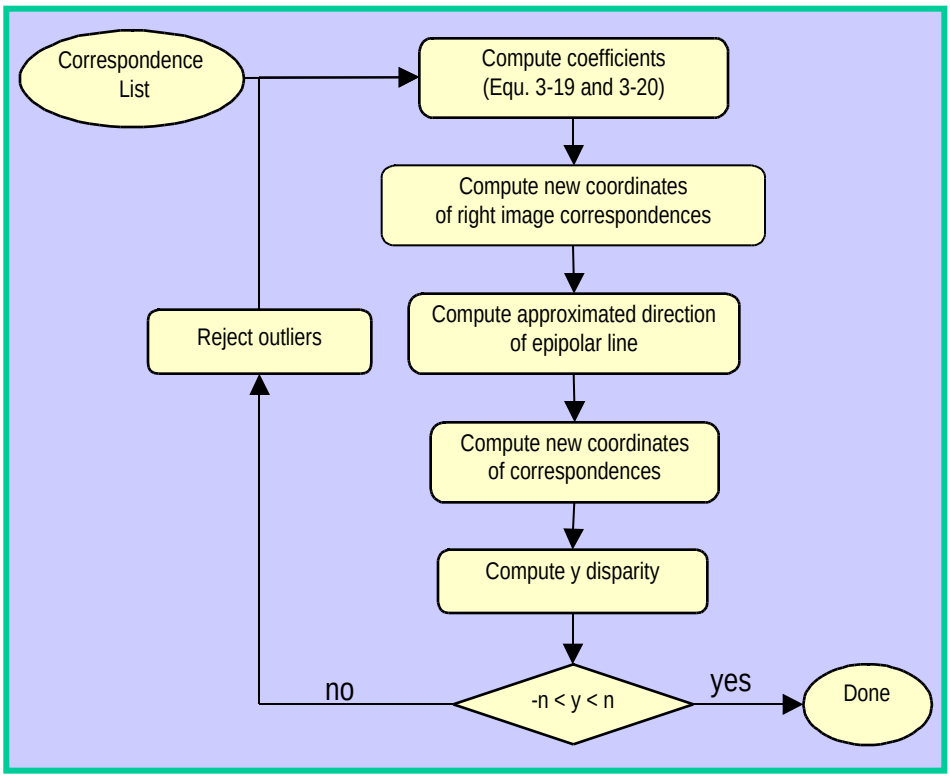

Figure 3-21. Flowchart of the algorithm of outlier detection and elimination

The regression diagnostic approach [Shapiro and Brady, 1995; Barnett and Lewis, 1984] is one of two main approaches (the other is robust regression) to solving the outlier problem. The classical diagnostic methods for outlier rejection compute an initial estimate over the full dataset; next, the residual for each data point is determined; and, finally, all points with residuals exceeding a predetermined threshold value (which is based on a chosen confidence level and a statistical-noise model) are rejected, the entire process being repeated in an iterative process as required. These methods are reliable when the outlier subset is relatively small and when outliers do not distribute far away from the valid data. These methods also

underlie a certain data distribution assumption (e.g. a certain statistical noise model, Gaussian Distribution).

In this research, a new algorithm is developed to reject outliers from a set of correspondences. Because outliers constitute only a small percentage of the data among correspondences after evaluating the best match, and their maximum deviation is bounded in the image tile size, in this new algorithm, a regression diagnostics approach is used to solve this outlier problem in an iterative process. However, neither a data distribution assumption nor an underlying statistical noise model is needed. The regression diagnostics is based on the vertical (Y) disparity.

As described in Chapter 2.2.3, the epipolarity for the linear pushbroom camera is an epipolar curve, and it is possible to approximate the epipolar curve by a straight line over a small range [Kim, 2000]. After the affine transformation along the epipolar line, the vertical (Y) disparities of correspondences must fall between $-n$ to n, where n is the number of pixels; n is set based on the epipolar curve and image size. To simplify the problem, n is set to be 3 for this prototype system. The existience of outliers will affect the evaluation of the angle of epipoplar lines (Equation 3-25), which causes an incorrect epipolar image re-sampling and generates large Y disparities of the correspondences between the epipolar stereo images. Once the outliers have been eliminated, the Y disparities will fall between $-n$ to n. Thus, the Y disparity of the correspondences between epipolar stereo images is chosen as the regression diagnostic. Because the outlier subset is small, outliers always have larger Y disparities than do correct correspondences. Therefore, the

correspondence with the largest *Y* disparity is rejected. The process of rejecting false matches is done by an iterative process that, for each loop, computes an estimate of the epipolar re-sampling over all correspondences; it then checks vertical disparities of all correspondences between epipolar stereo images; and, finally, deletes a correspondence outlier.

The following steps summarize the use of the algorithm in an iterative process, as illustrated in Figure 3-21.

1. Compute the affine transformation coefficients (Equations 3-19 and 3-20) for the affine transformation of the right image to the left image.
2. Compute the new affine coordinates of right image points of correspondence after transformation according to Equations 3-19 and 3-20.
3. Compute the approximated direction of the epipolar line (Equation 3-25).
4. Compute the affine coordinates of the left and right image points of correspondences after rotating according to affine rotating equations.
5. Compute the *Y* disparity for each correspondence.
6. If the absolute values of some *Y* disparities are larger than *n* (regression diagnostics), reject the outlier. It updates the correspondence list by eliminating the correspondence with the largest *Y* disparity. Repeat step 1.
7. Otherwise, the iteration stops.

As shown in Figure 3-21, the input to this process is a correspondence list. Outputs of this process are the coefficients (Equations 3-19, 3-10, and 3-25) for epipolar image re-sampling and an updated correspondence list.

3.4 3D Image Generation: Epipolar Image Re-sampling and Anaglyph Image Formation

Since the coefficients of Equations 3-19, 3-20, and 3-25 have been determined during the process of eliminating outliers, the general procedure of epipolar image re-sampling may be described in the following simplified form:

- Register the right image with the left image according to Equations 3-19 and 3-20; and then
- Rotate both the left and transformed right images by the angle of the epipolar line.

In all stages processed so far, the epipolar images as stereo pairs are generated from stereo images using the 2D affine orientation model. The next step is to form colour anaglyph images, which superimposes the epipolar stereo pair by merging green and blue bands of the left image and the red band of the right image together to form a colour composite image (anaglyph image). During this process, the horizontal direction superimposition is considered to create a consistent and easily perceived impression of depth by the author. As described in chapter 2.1.2, depth perception is induced by X parallax. However, there is an asymmetric depth perception between the crossed disparity and the uncrossed disparity. Also, the depth perception with crossed disparity has closer predictions than that with

uncrossed disparity in most cases. Moreover, one of two depth perception directions (crossed and uncrossed disparities) is impaired to stereo-anomalous observers. Therefore, the criterion of superimposing in the horizontal direction is to compel the *X* disparity of the correspondence with the smallest *X* disparity to be zero.

Unfortunately, during the above two processes - that is, in both epipolar image re-sampling and anaglyph formation - the x and y coordinates are not always computed as exact integer values and thus may not fall exactly within the centre of the pixel. Therefore, resampling is needed to decide the exact value for each pixel. Three resampling approaches were incorporated into the epipoplar image re-sampling and anaglyph formation processes developed for this research: nearest neighbour resampling, bilinear interpolation resampling, and cubic convolution interpolation. An option is provided for users to choose a resampling method in this system (see Figure 3-9).

3.5 3D Web Visualization

3.5.1 User Interface for 3D Image Visualization

In this system, two kinds of user interfaces are provided to allow users to interactively view 3D images: one for local 3D visualization (Figure 3-22) and the other for 3D visualization through Web browsers (which will be described in section 3.5.2). For the first case, a GUI in the form of a 3D image viewer is implemented in Java programming language; it consists of an extension of the JAI API to provide the interface for image visualization. Both of these components provide the following services:

1) Showing two kinds of images - colour 2D and 3D - on a viewing window;
2) Allowing a user to zoom in on the image in the viewing window by mouse clicking the zoom in button or to zoom out by mouse clicking the zoom out button;
3) Allowing a user to pan the image in the viewing window;
4) Switching 2D and 3D image viewing in the viewing window; and
5) Automatically or interactively refining the 3D effect in the viewing window.

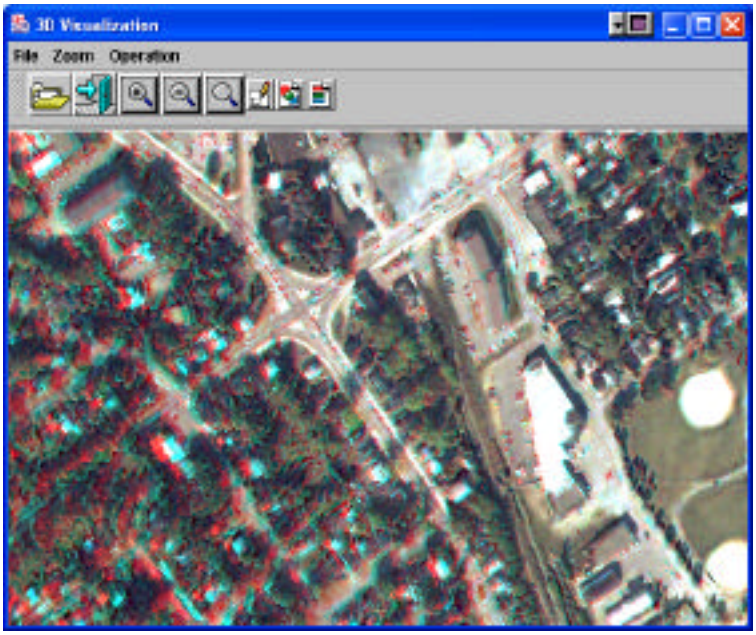

Figure 3-22. User interfaces for 3D visualization (local 3D viewer)

A key issue is the efficiency of the automatic parallax adjustment for an optimal 3D display to meet the real time requirement. In this research, a Quadtree-

based method is proposed to solve this issue for more efficient 3D effect refining. To build a Quadtree, correspondence information of quadrants is required. The stereo match process is used to obtain correspondence information. The Quadtree index structure, the stereo match process, and the 3D refining method used in this research are described in the next three sections.

3.5.2 Quadtree

The Quadtree structure has been widely used in the fields of image processing [Fazekas and Santa, 2000; Yang et al., 2000; Manouvrier et al., 2002], computer graphics [Samet and Webber, 1988], GIS [Shaffer et al., 1990; Wang and Armstrong, 2003], image databases [El-Qawasmeh, 2003], and image compression [Wang et al., 2000; Chung and Tseng, 2001]. A Quadtree is a member of a class of hierarchical structure based on the recursive division of the image in four disjoint quadrants, and is well suited to applications that involve searches [Samet, 1990]. Quadtrees can be used to represent different types of data; that is, point data, areas, volumes, and so on. The most known and studied Quadtree approach is the region Quadtree. It partitions an image into four equal-size quadrants by means of recursive successive subdivision. The partition will be stopped once a given criterion is met, and so the resolution of decomposition (the number of times it is decomposed) can be changed according to the properties of the input data.

Figure 3-23 presents an example of a binary image (2^3 by 2^3 pixels) and its regional Quadtree representation. The root of the Quadtree corresponds to the entire binary image. The descendant nodes of a node (parent) represent quadrants (labeled

NW, NE, SW, and SE) of the region represented by that parent node. A leaf node corresponds to a homogeneous image quadrant (all white or black pixels) for which no further subdivision is necessary. An internal node contains both black and white pixels. As shown in Figure 3-23, the leaf nodes of a Quadtree (unbalanced tree) are not at the same level in general.

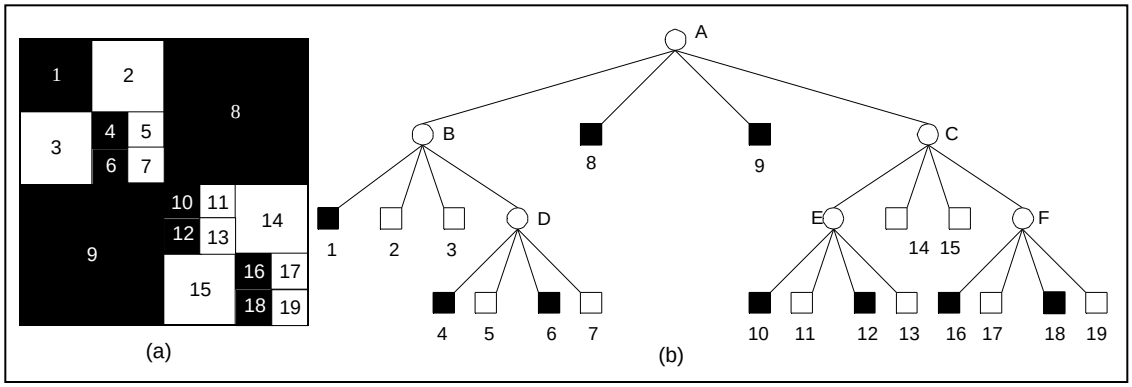

Figure 3-23. Examples of (a) a binary image
and (b) the corresponding Quadtree representation

3.5.3 Stereo Match

Refining the 3D effect for a particular area requires the identification of correspondences with the smallest *X* disparity in the refining area. Once the epipolar images of the stereo pair have been established, epipolar stereo images can be used to establish more new correspondences using both feature-based and area-based methods that consider the epipolar constraint.

In this research, some modifications have been made to the matching approach described in section 3.3. The first is image tiling, while the second considers the epipolar constraint. As illustrated as Figure 3-24, decomposition of

image tiling has the property that, at each subdivision stage, the image is subdivided into four equal-sized parts. This allows us to use the Quadtree index structure to search the correspondence information in each tile area for later refinement of the 3D effect. The level *l* of decomposition depends on the sizes of the image and 3D viewer, defined by this research as follows:

$$level = Math.\max\left(\left\lfloor \log_2 \left(w_1 / w_2 \right) \right\rfloor, \left\lfloor \log_2 \left(h_1 / h_2 \right) \right\rfloor\right) + 1 \qquad 3\text{-}30$$

where w_1 and h_1 represent the image width and height, respectively; and w_2 and h_2 represent the viewer width and height, respectively.

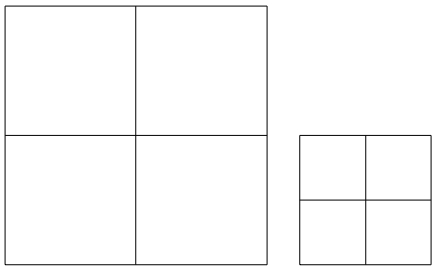

Figure 3-24. Decomposition of Image Tiling

After feature points are extracted, for a feature point of the source (left) image, a local support region is defined to search the possible matches of the target (right) image near the epipolar line. The thickness of the local support region can be set based on the epipolar curve and image size. In this system, the thickness is defined as 3 to simplify the problem. To select the most consistent matches, the strength of the match and uniqueness constraint are applied. At this time, for each

image tile, correspondences with the smallest disparity are selected and saved to a project file for further processing to refine the 3D effect.

3.5.4 Refining 3D Effect

Depth perception in binocular vision is caused by horizontal parallax. However, excessive horizontal parallax will cause depth distortion and stereo fusion problems, and so the parallax should be maintained in a limited range for an acceptable stereo fusion (e.g., less than 30 arcmin) [Patterson, 1992; Carr, 1993; Patterson, 1997]. Since satellite images cover a large area, as illustrated in the left image of Figure 3-25, some areas with high elevation have larger horizontal parallaxes. 3D refining automatically reduces the horizontal parallax to maintain an acceptable parallax (see Figure 3-25, right image). The criterion for refining the 3D effect in a viewing area is similar to that for anaglyph forming, which compels the X disparity of the correspondence with the smallest X disparity to be zero within the viewing area.

Figure 3-25. An example of horizontal parallax, left image with larger horizontal parallax, and right image with small horizontal parallax

In this research, a Quadtree-based method is proposed to allow for more efficient 3D effect refining, and a modified bucket PR Quadtree is developed. Initially, the bucket methods were designed to ensure efficient access to disk data. A bucket PR Quadtree is a region-based Quadtree in which regions are decomposed until they do not contain more data points than the bucket capacity.

Figure 3-26 presents the image with correspondence information and the representation of the modified bucket PR Quadtree in this research. The Quadtree used in this system stores information about correspondences with the smallest *X* disparity in each quadrant of the stereo images. In the modified bucket PR Quadtree, there is no limitation of bucket capacity. The decomposition of the Quadtree is based on the successive subdivision of an image (in this system, the left image of a stereo image pair is used as a split image) into four near equal-size quadrants (the image

size is not always 2^n). The resolution of the decomposition is given by Equation 3-30, and the Quadtree represents a complete tree. The root of the Quadtree corresponds to the whole image. The descendant sub-nodes of the root correspond to quadrants within the image.

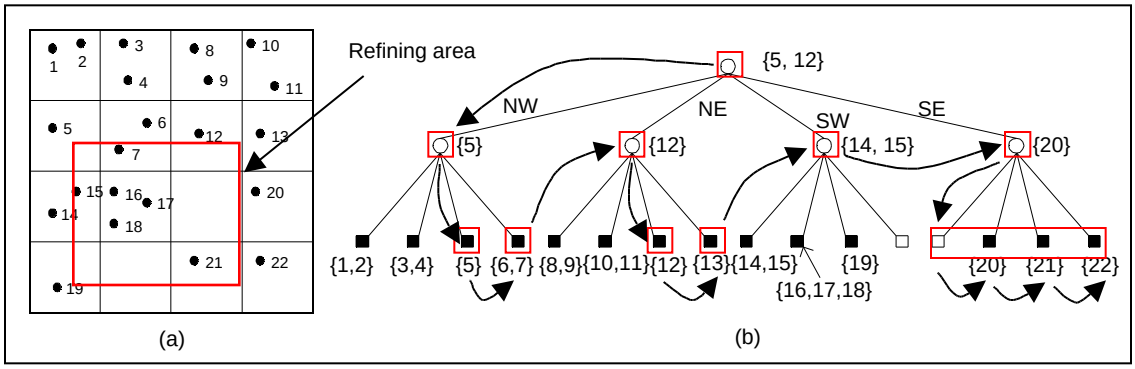

Figure 3-26. Quadtree method for 3D parallax refinement
(a) Decomposition of an image, and (b) Quadtree representation of the image.
· -- Match, number (e.g. 1, 2) – the order of Matches, { } – Vector which contains the Matches

Each node of the modified bucket PR Quadtree contains five fields. Four fields point to the node's four children corresponding to the four quadrants. One field, using a Vector, stores information about the object *Matches*. Each *Match* contains information about the disparity (value) and the coordinates of the correspondences with the smallest horizontal parallax within this quadrant.

The building of the modified bucket PR Quadtree uses the correspondence information obtained in the stereo match process. The algorithm is as follows:

(1) Initialize the Quadtree root to the entire image with a null Vector and four null children.

(2) Build an empty complete Quadtree with the depth *level* computed by Equation 3-28. Each node of the Quadtree has a null Vector.

(3) Read correspondence information within one image tile from the project file (a data stream for the Web application).

(4) For each correspondence, construct a *Match* object and store it in a Vector *V*.

(5) Insert the Vector into the Quadtree.

- Processing starts from the root upon reaching the bottom (leaf) of the Quadtree.
- At each internal node *R*, two kinds of comparisons are made between the Vector *V'* of node *R* and the Vector *V*:
  - One comparison is based on the values of their *X* disparity. If the *X* disparity of matches of *V'* is larger than that of *V*, then *V'* is replaced by *V*. If they are equal, all the matches of *V* are appended to *V'*. Then, skip to step 6. Otherwise, continue to the comparison below.
  - The second comparison is based on the *x* and *y* coordinates. This comparison returns the quadrant (child's node *P*) of the Quadtree rooted at the node R in which the Vector *V* belongs, and the appropriate subtree is chosen for the next test.
- When Vector *V* reaches the bottom of the Quadtree, the location where it is to be finally inserted is identified.

(6) If the end of the file (or data stream) is encountered, then stop. Otherwise, repeat step 3.

Now, 3D refining is converted to a range search in a Quadtree to find the minimum disparity within the refining area so that it improves the efficiency of the system. To refine the 3D image effect, we simply search for information about those quadrants that are crossed with the refining area (Figure 3-26). In the searching of the Quadtree *Q*, the root of *Q* is visited first, and then part, or all, of the subtrees rooted as its children are traversed recursively. When a node *N* is visited, it checks if there are *Matches* stored at *N* within the refining area (see Figure 3-26). If *Matches* are found, then traversing the subtrees of *N* ceases. Otherwise, the algorithm continues to examine which quadrants (substrees) of *N* are crossed with the refining area, and then returns minimum (disparity values of the determined quadrants). The disparity values are obtained by visiting the determined quadrants. In Figure 3-26, the arrows show the traversal of the search process for the minimum disparity value.

The searching process is summarized as follows:

Algorithm searchMinDisparity (*Q*, *Q*.root, *area*)
    If there are *Matches* stored at *Q*.root within the refining *area*
        Return disparity value of the Matches
    Otherwise {
        Find the children (quadrants) of *Q*.root cross the refining area
        Return Min(for each child *w*, serchMinDisparity(*Q*, *w*, *area*))
    }

Because the number of Matches stored in each node are limited, the worst case computation time of the search is $O(n)$, where n is the depth of the Quadtree.

3.5.5 Establishing Web-based Interactive 3D Visualization

This portion of research aims to develop a Web-based client/server architectural solution for 3D visualization through Web Browsers. At this time, 3D Web visualization only presents the 3D image online, and there is not much processing management (no business logic computation) required in the system. In addition, there is no data other than image data. Therefore, these image data (image files) are directly stored on the Web server.

Figure 3-27 shows an overview of the architecture of the Web-based client/server model used in this research. With this architecture, a platform-independent system is implemented. Any user who links to the Internet and has a Web browser can use this system. Moreover, it is easy to deploy the Web contents (files) to the Web (production) server. For example, one can use FTP or other file-transfer software of her/his choice to transfer the Web page files to the production server.

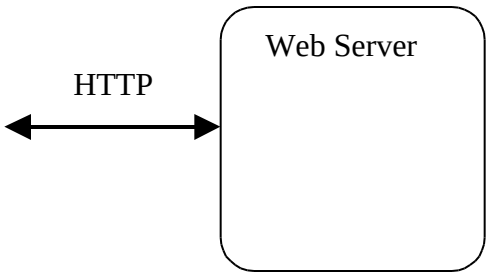

Figure 3-27. Architecture of the Web-based client/server model of the W3VPS

3.5.5.1 The Client-tier Components

A 3D image display requires a rich interface, which is tied to the behaviour of the application's client tier. A client makes requests to the server and presents the outcomes to the user. Thus, a client configuration must be well chosen so that it addresses the requirements of the application and provides users with a robust interface. Web clients use HTTP as the transport protocol and usually operate inside a browser. Browsers are widely available because users are familiar with them, and there are no issues with deployment. Applets are Java-based GUI components executing in a Web browser that are used to enhance the browsing interface. The Web clients can use any Java-capable Web browser to run the Java-based client components. In this research, the Web client configuration combines both stand-alone Web browsers and applets. Figure 3-28 shows the implementation of a client applet, "3DViewApplet," which is implemented to provide the user interface for the 3D visualization in the client tier. It provides the services described in section 3.5.1. The solution of 3D effect refining is the same as that presented in section 3.5.4.

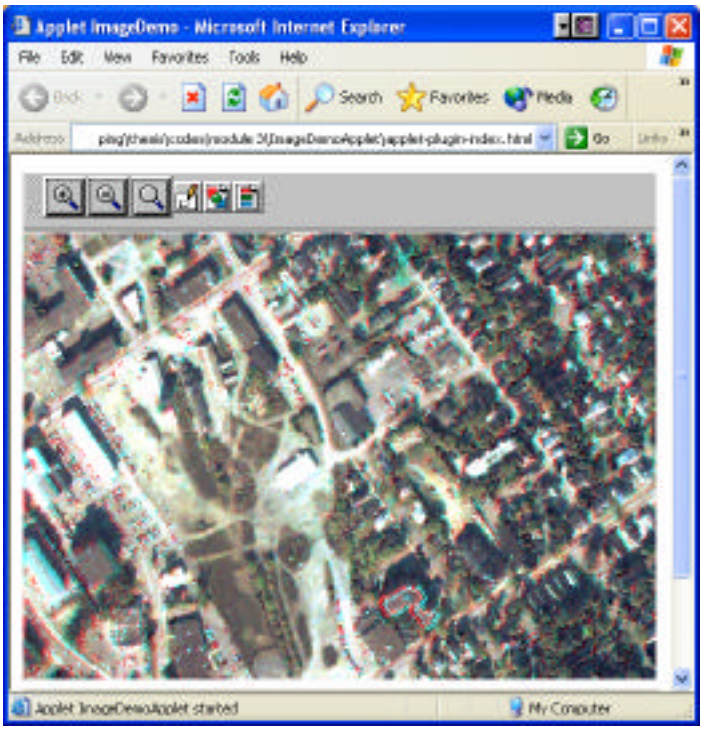

Figure 3-28. User interface of the Web-based 3D visualization applet

3.5.5.2 Applet -Web Server Communication

A Web server is used in this research to store Web documents (image files) and to make them available to the online user community. The communication between an applet and the Web server will be established by an HTTP connection to allow the applet to connect to the Web server machine.

Most of the Internet users access the Internet through relatively low speed modems, but the size of satellite images containing both 2D and 3D information is relatively large in this research. Therefore, a significant problem exists in terms of bandwidth constraints. To overcome this problem, a technique is used in the W3VPS that allows the division of high-resolution images including four-colour

channels into bite-size tiles and to deliver “pixels on demand”. In this case, data transmission is done by downloading only the required files on demand, which allows for high-resolution images over all bandwidth environments.

An image-tiling component is implemented to split the original larger images into the image tiles. Thus, by limiting the amount of data that needs to be exchanged, data delivery on a “pixel on demand” basis between the Applet and Web server becomes feasible. Figure 3-29 shows an example of the image tiling that partitions the original image into rectangular, non-overlapping, blocks with the same dimensions (100 by 100 pixels in the W3VPS) except for those on the boundaries of the image. Then, the tiling process stores the image tiles onto the hard disc in .jpg format. By default, the output image tile files are named by their tiling orders in the row and column, prefixed with an "im". For example, see Figure 3-29 (right image), where the name of the image tile of the third row and third column is “im0303”. Naming the image tiles in this representative way facilitates the Web client’s (Applet’s) access and retrieval of the data on a “pixel on demand” basis from the Web server.

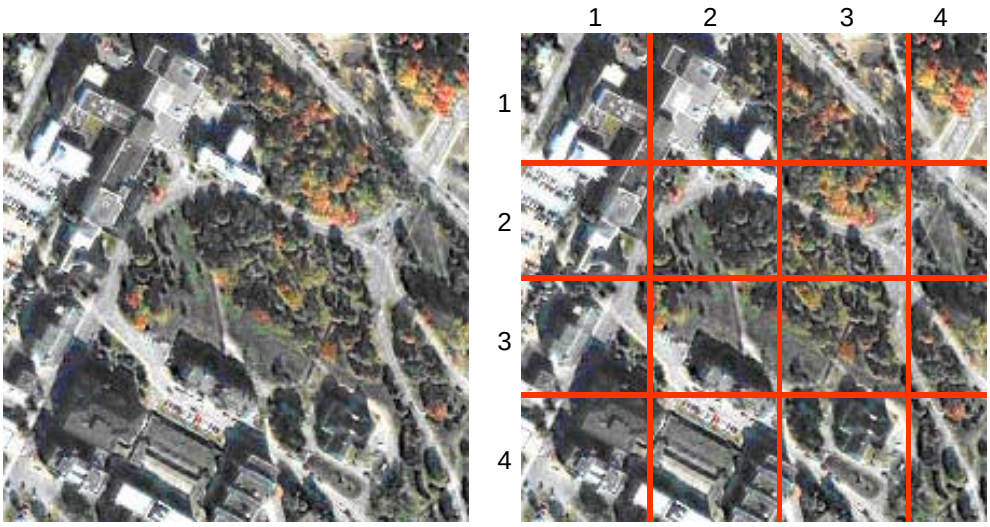

Figure 3-29 Results of the image tiling. Left: original image; Right: image tiles

Communication between the Applet and Web server occurs in the following way:

The Applet sends a request to the Web server (e.g. zoom in image).

(1) The Applet sends a request to the web server (e.g. zoom in image).

(2) Web server sends the response (images) in HTML form to the Web client by HTTP protocol.

(3) The Applet displays the transmitted image data on the screen.

Chapter 4

Testing and Evaluation

The effectiveness of the proposed solution was evaluated by testing each component of the W3VPS. The tests performed on the system included image-matching, 3D image generation, and 3D Web visualization. This chapter first describes the computing environment which was used to develop and test the W3VPS; then presents the component test results and their analyses; and finally, summarizes the test results to evaluate overall system performance.

4.1 Computing Environment: Hardware and Software Metrics

The software and hardware components used to develop and test the W3VPS are summarized in Table 4.1 and Table 4.2, respectively.

Table 4.1. Software for system development and test

Component	Software Used
Image Processing Components	Java(TM) 2 SDK, Standard Edition Version 1.4.2 Java Advanced Imaging (JAI) API, Version 1.1.1_01
Web Server	Apache/2.0.42 Server
Client Browser	Internet explorer 6.0

Table 4.2. Hardware and operating system for system development and test

Component	Make/Model	CPU	Memory	Operating System
Image Processing Components	Dell OPTIPLEX GX260	Pentium(R) 4 1.80GHz	1.00 GB	Microsoft Windows XP Professional
Web Server	Dell OPTIPLEX GX260	Pentium(R) 4 1.80GHz	1.00 GB	Microsoft Windows XP Professional

As listed in Table 4.1, the W3VPS developed for this research used Java (JDK 1.4) as a development platform and an extension of the JAI API 1.1.1_01 for 3D image generation and 3D Web visualization. Both JDK 1.4 and JAI API 1.1.1_01 are in the public domain and available free of charge via the Internet (SUN 2003(a) and SUN 2003(b)); they were installed on one machine, *studio*, in the Geographic Information Group Lab at UNB.

Although many Web servers exist, provided by various vendors, the Apache HTTP Server, is the most popular Web server on the Internet and is freely downloadable. Apache [2003] provides details about the features of the Apache HTTP Server and its configuration. In this research, Apache/2.0.42 Server was installed on *studio*. Also, Microsoft Internet Explorer 6.0 as a widely used browser was chosen as the client browser for the test.

4.2 Testing Results and Analysis

The W3VPS for high-resolution satellite stereo images was tested for both 3D image generation and 3D Web visualization. The testing was based on the following three components: image-matching, 3D image generation, and 3D Web

visualization. They are described in the following sections including the descriptions of the functionalities tested, followed by the testing methods and result analysis for each.

4.2.1 Image-matching

The proposed image-matching algorithm developed in this research was tested; see section 3.3 for the details on the image-matching algorithm.

The quality of the matches is crucial to 3D image generation, since mismatches affect the accuracy of epipolar image re-sampling, thereby detracting from the visual clarity of the 3D image. Because the distribution of the correspondences affects the coefficient evaluations (Equations 3-19, 3-10, and 3-24) required for epipolar image re-sampling in 3D image generation, the results of the distribution of the correspondences were tested. The evaluation of the image-matching algorithm was, therefore, based on: (1) the quality of the correspondences; (2) the distribution of the correspondences; and (3) the efficiency of finding the correspondences.

As discussed in section 3.2, the city of Fredericton is the study area. Original stereo images (with the different sizes) of around 3000 by 4500 pixels were tested in the experiment. All parameters used in testing the proposed algorithm were set according to the test results of the image-matching Component 1 (see section 3.3, the test component). Therefore, the Gaussian standard deviation (σ) is defined as 1; based on this, the Gaussian filter is used to generate the corner images. The size of local extrema search windows is 13 by 13 pixels. The sizes of the correlation search

window and the template are set at 100 by 100 pixels and 9 by 9 pixels, respectively. The threshold of the correlation coefficient is defined as 0.9.

The next three sections describe the testing of the quality of the correspondences, the distribution of the correspondences, and the efficiency of the method of finding them.

4.2.1.1 Quality of the Correspondences

The accuracy of the results of the proposed image-matching algorithm was tested. As described in section 3.3, the new image-matching algorithm consists of four stages: (1) feature point extraction; (2) searching for candidate matches; (3) evaluating the best match; and finally, (4) elimination of outliers (see Figure 3-4 for an overview of the new image-matching algorithm). In this new algorithm, a modified SM (strength of match) method based on Zhang's SM [Zhang et al., 1995], was developed for evaluating the best match. Zhang's was also implemented to compare the performance of the two SM methods. The new image-matching algorithm also included a routine for eliminating outliers, referred to herein as the 'eliminating outlier algorithm (EOA)'. Thus, the performances of two versions of image-matching process were compared: one using Zhang's SM and the other using my modified SM (proposed image-matching algorithm). This testing was done manually.

Table 4.3 shows the results of the SM methods for both image-matching processes. Table 4.4 shows the results of the EOA which took the data produced by Zhang's and modified SM as input, separately. In Table 4.3, the *Total/bad* shows

the numbers of total and bad matches; only false matches are classed as bad matches. The *residual* includes both false matches and corner localizing errors. To specify the residual of the matches, *RMS* in pixels is counted in *x*, *y*, and *xy* directions separately. The *total* subheading specifies the total number of RMS of all matches, while *average* specifies the total residuals against the number of matches (average = total / number of matches). In Table 4.4, the *detected* represents the number of outliers detected by the EOA Similarly, the total and average RMS in pixels in *x, y,* and *xy* directions are listed under the row *residual*. The row labelled, *No. of Iterations* specifies the number of iterations of the algorithm performed during the process of eliminating outliers.

In Table 4.3, in the *Zhang's SM* column, the results obtained from Zhang's SM method, while the results given in the modified SM column obtained from the modified SM method. In Table 4.4, in the second column, the results were obtained from the EOA using data produced by Zhang's SM method while the results given in the third column were obtained from the EOA using data produced by the modified SM method.

Table 4.3. Results of the Evaluating the best match algorithm (Zhang's SM method and modified SM method separately)

Evaluating the Best Match			Zhang's SM	Modified SM
Total/bad			143/21	143/2
Residual (RMS in pixels)	x	Total	213	26
		Average	1.4859	0.1818
	y	Total	115	11
		Average	0.8042	0.0769
	xy	Total	285.32	36.04
		Average	1.9952	0.2520

Table 4.4. Results of the EOA

Eliminating Outlier Algorithm			Using data produced by Zhang's SM (21 bad matches) as input	Using data produced by modified SM (2 bad matches) as input
Detected			18	1
Residual (RMS in pixels)	x	Total	33	15
		Average	0.2661	0.1056
	y	Total	23	10
		Average	0.1854	0.0704
	xy	Total	56	25
		Average	0.4516	0.1760
No. of Iterations			19	2

Table 4.4 shows that good results have been obtained using the proposed image-matching algorithm. For example,

Average residual RMS x = 0.1056 pixels and

Average residual RMS y = 0.0704 pixels

Explanations for evaluating the best match:

Table 4.3 clearly shows that the modified SM method is efficient and superior to Zhang's SM in terms of the quality of matches that it returns, witness the following comparisons:

- Average residual RMS x is reduced from 1.4859 pixels to 0.1818 pixels.
- Average residual RMS y is reduced from 0.8042 pixels to 0.0769 pixels.
- Bad matches are reduced from 21 to 2.

Finally, as shown in Table 4.4, the results of the EOA using the correspondence data produced by the modified SM method were compared with

those using data produced by Zhang's SM method. The comparison confirmed that the modified SM method is more effective than Zhang's, as shown below:

- Average residual RMS x is reduced from 0.2661 pixels to 0.1056 pixels.
- Average residual RMS y is reduced from 0.1854 pixels to 0.0704 pixels.

In Zhang's SM, the contributions of candidate matches in the neighbourhood are weighted by their distances from the match. This means that a close candidate match gives more support to the match being measured than a distant one. Thus, when most candidate matches of tiled stereo image pairs join closely together, false matches may result. In the modified SM approach, the weights for all candidate matches are the same. This still satisfies the rule that an affine approximation is reasonable only for a small area because the sizes of tiled images (candidate match neighbourhood) are small.

***Explanations for eliminating outliers*:**

In Tables 4.3 and 4.4, the average *x* and *y* residuals before and after rejecting the outliers show the improvement achieved by incorporating the outlier detection module. The detected numbers also demonstrate that the EOA rejects the majority of outliers. The identified outliers were also inspected manually and confirmed as incorrect matches. Although not all of the outliers were removed by either approach, those few that lie along the epipolar lines have a negligible effect on epipolar image re-sampling.

The other intuitive error measures used to evaluate the quality of the correspondences (matches) include the sum, average, and maximum value of the *Y*

disparities (parallax) of the matches after epipolar image re-sampling. As shown in Table 4.5 and Figures 4.1 to 4.3, these measures demonstrate the results obtained from the EOA. Because Zhang's SM produced more *bad matches* (21 bad matches over 143 total matches; see Table 4.3), the comparison of Zhang's SM method in the evaluation of the EOA as the immediate computational precedent is appropriate.

In Table 4.5, the results given by the *before rejecting outliers* column were obtained directly from Zhang's SM method. The results yielded by the *after rejecting outliers* column were obtained from the EOA using data produced by Zhang's SM method.

Table 4.5. Comparison of *Y* disparity and rotation angle before and after the outlier rejection

	Before Rejecting Outliers	**After Rejecting Outliers**
Sum of *Y* disparity (pixel)	448.39	73.58
Max *Y* disparity (pixel)	29.40	2.71
Average *Y* disparity (pixel)	3.80	0.73
Rotation Angle (arc)	0.5856	0.4604

Table 4.5 compares the results before and after the outlier rejection algorithm is applied based on the following four aspects: *sum of Y disparity*, *max Y disparity*, *average Y disparity*, and *rotation angle*. The *sum of Y disparity* represents the sum of *Y* disparities of all matches after epipolar image re-sampling. *Max Y disparity* represents the maximum value of *Y* disparities of all matches after epipolar image

re-sampling. *Average Y disparity* represents the *sum of Y disparity* against the number of the matches. *Rotation angle* represents the angle of the epipolar line.

As shown in Table 4.5, *Sum of Y disparity*, *Max Y disparity*, and *Average Y disparity* in the *After Rejection Outliers* column were greatly reduced from their corresponding values in the *Before Rejecting Outliers* column. The *Rotation Angle* was considerably altered after outlier rejection. On this basis, the EOA appears to be effective in the rejection of outliers.

Figures 4.1 to 4.3 show the intermediate results during the execution of the EOA based on three aspects: maximum *Y* disparity, average *Y* disparity, and rotation angle. In these three figures, the *x*-coordinate represents the number of loop times.

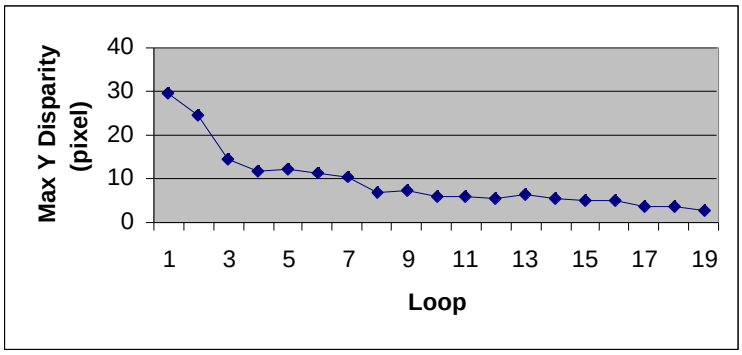

Figure 4-1. Maximum *Y* disparity comparison of the loops of the EOA (using 21 bad matches over 143 total matches as input)

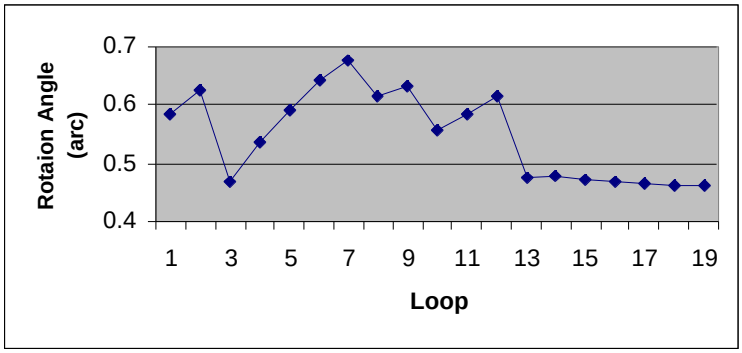

Figure 4-2. Rotation angle comparison of the loops of the EOA (using 21 bad matches over 143 total matches as input)

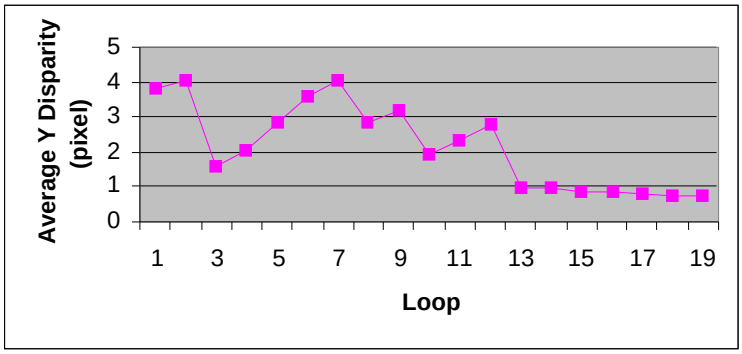

Figure 4-3. Average *Y* disparity comparison of the loops of the EOA (using 21 bad matches over 143 total matches as input)

In Figure 4.1, the *y*-coordinate represents maximum *Y* disparity. The Maximum *Y* disparity was reduced with successive iterations. In Figure 4.2, the *y*-coordinate represents rotation angle. The rotation angle fluctuates before loop 13 because mismatches exist with large residuals, and these mismatches affect and control the result of the rotation angle. Even at that stage, the rotation angle is always greater than the rotation angle for loop 13. Rotation angle reduces smoothly subsequent to loop 13. In Figure 4.3, the *y*-coordinate represents the average *Y*

disparity. The average *Y* disparity fluctuates similarly to that in Figure 4.2 because the *Y* disparities of matches after epipolar image re-sampling are calculated based on the angle of the epipolar line. Thus, testing results indicate that the EOA progressively improves the accuracy of the results.

4.2.1.2 Distribution of the Correspondences

The distribution of correspondences over the stereo image pairs was tested in this research. Investigating this aspect of the results is important because the distribution of the correspondences directly affects the stability and accuracy of epipolar image re-sampling.

Drawing the results (a set of correspondences) of the proposed image-matching algorithm over the stereo image pair provided another means of testing the result of the distribution of correspondences. In this research, only the points of the correspondences in the left image were used as test data. Figure 4-4 shows the results of the distribution of the correspondences obtained by the new image-matching algorithm. The *x* direction represents the column of an image, while the *y* direction represents the row of an image. Dots represent the points of the correspondences in one of the stereo images.

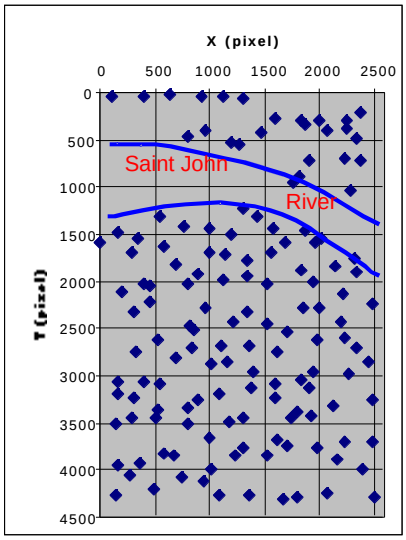

Figure 4-4. Distribution over the left image of the correspondences

Figure 4.4 shows the good result that the points of the correspondences are evenly distributed over the whole image with the exception of the river area that is a homogenous area and in which no correspondences can be found. Currently, there are no image-matching approaches to effectively find the correspondences of homogeneous areas. Thus, the new image-matching algorithm, using the image tiling technique, results in the effective distribution of the correspondences.

4.2.1.3 Efficiency of Finding the Correspondences

The processing time for image-matching was tested in this research as an additional means of evaluating of the effectiveness of the image-matching algorithm.

In testing system performance, each tiled stereo image pair as well as the whole stereo image pair was tested in each of the six stages: tiling, corner image

generation, corner extraction, correlation, best match, and outlier detection. Table 4.6 shows CPU time in milliseconds spent in each step for each tiled stereo image pair and the whole stereo image pair. In this table, *Tiling* represents the image tiling process; *corner image* represents the corner image generation process; *corner extraction* represents the corner detection process; *best match* represents the process of evaluating the best match; *correlation 1* represents the correlation process with neighbourhood disparity information; *outlier detection* represents the eliminating outlier process.. *Total* represents the total processing time and was counted at the end of image-matching test. In Table 4.7, *correlation 2* represents the correlation process without neighbourhood disparity information.

As described in section 4.2.1, the sizes of the original stereo images are around 3000 by 4500 pixels. The image tile size, except for the boundary tiles, is defined as 200 by 300 pixels, or 220 by 320 pixels, for the left image or right image respectively; there exist 168 tiled stereo image pairs.

Table 4.6. Computing time in each stage of the image-matching process: CPU time in milliseconds

Matching stages	**Each tiled image pair** (average 168 tiled pairs)	**Whole image pair**
Tiling	720.19	130992
Corner image	0.28	48
Corner extraction	1139.1	191370
Correlation 1	495.61	83263
Best match	1.75	295
Outlier detection		31.42
Total	**2356.95**	**395999.42**

Table 4.7. Computing time in the correlation without the neighbourhood disparity information: CPU time in milliseconds

Matching stages	**Each tiled image pair** (average 168 tiled pairs)	**Whole image pair**
Correlation 2	1075.09	180616

The time values shown in Tables 4.6 and 4.7 indicates that the computing time for correlation is reduced significantly – that is, by over 50% (see row *correlation 1* in Table 4.6 and *correlation 2* in Table 4.7) - when disparity information obtained from neighbouring tiled stereo image pairs was used. The time taken for correlation depends on tile scene complexity (the number of corner points) and neighbourhood disparity information according to the observation of both time difference and scene content, between each tiled stereo pair. The times attributed to image tiling, corner extraction, and best match location are essentially constant throughout the process of iterations. In general, the most time-consuming steps are image tiling, corner extraction, and correlation.

In this research, a new robust approach to image-matching by exploiting the epipolar constraint has been proposed, developed, and tested. The principal innovation brought by the new approach is in using certain heuristic techniques (corner detection and correlation methods in this implementation) to find initial matches, and then to use a modified SM approach to locate the best match within a tiled stereo image pair; finally, an outlier detection module (the EOA) uses a robust technique – the regression diagnostic approach – to discard outliers.

The testing was completed, and the results obtained in this research indicate that the new approach works effectively by comparing with the matching results shown in the literature [Shapiro and Brady, 1995; Zhang et al., 1995; Zhou and Shi, 2002].

4.2.2 3D Image Generation

3D image generation was tested to evaluate the results of anaglyph images for the purpose of 3D image accuracy evaluation. In general, epipolar images are used for automatic DEM extraction or 3D stereo viewing to minimize registration error in the *y*-direction. Also, as described in section 2.1.3, the *Y* parallax should be minimized for correct depth judgements. The RMS of the *Y*parallax is, therefore, used as a measurement to evaluate the quality of the 3D image for 3D viewing.

For verification purposes, 68 checkpoints, distributed evenly over the entire stereo image pair, were measured manually, the results of which are shown in Table 4.6. In this table, *RMS of y-parallax* represents the total number of *Y* parallaxes, in pixels, associated with the 68 checkpoints; *RMS (average) of y-parallax* represents the total *RMS of y-parallax* against the number of checkpoints in pixels; and *Maximum y-parallax* represents the maximum y-parallax among the checkpoints, again, in pixels.

Table 4.6. Testing results of the *Y* parallax of an anaglyph image

RMS of y-parallax (pixel)	**RMS (average) of y-parallax (pixel)**	**Maximum y-parallax (pixel)**
29	0.4264	2

Table 4.6 shows that accuracy of better than half a pixel was achieved and that the maximum *Y* parallax was 2 pixels under practical experimental conditions. The results obtained for this research indicate that the epipolar re-sampling approach, based on the 2D affine orientation model, can be effective for the mapping of the medium-sized areas (e.g., 4500 by 4500 pixels) of the high-resolution satellite stereo image pairs for the natural environments.

4.2.3 3D Web Visualization

3D Web visualization was tested to evaluate the performance of its architectural constructs upon implementation. Testing focused on evaluating the results of refining 3D image effects and the efficiency of functionalities, such as the zoom image, provided by the Web client.

All tests were conducted on the same machine, *Studio* (see table 4.2 for the characteristics of this machine), because of the limitation of the licence requirement to publish JAI contents on the Web. In this research, all tests were examined using the Internet Explorer 6.0 Web browser running on the same machine as the Web server. When commercial software, in the form of *Viewpoint*™, was used for image visualization, no significant differences were observed in the data transmission time of different machines in the UNB LAN. The 3D Web visualization uses technology similar to that of *Viewpoint*™ for data transmission. Thus, on the assumption that these machines faithfully represent those on other (typical) networks, we may infer that the results obtained from testing 3D Web visualization on *studio* are representative.

4.2.3.1 3D Image Refinement

To achieve acceptable stereo fusion in which there is no the process of binocular rivalry (see pages 16 and 17 for the details about binocular rivalry), the *X* parallax should be maintained within a limited range. As described in Chapter 3, the objective of 3D image refinement is to adjust the *X* parallax of the 3D viewing area of a 3D anaglyph image for optimal 3D display, in real time. In this research, the testing for 3D image refinement includes two parts: 3D effect and refining time. An original anaglyph image of approximately 2000 by 2350 pixels was tested, the viewing area of which was defined as 500 by 360 pixels.

3D effectiveness:

An optimal 3D effect facilitates depth perception of the viewing area without the depth distortion and stereo fusion problems brought about through previous approaches (Anaglyphs). Being essentially a subjective quantity, the quality of the 3D effect was tested by 10 observers with the different backgrounds (some of them have no 3D viewing experiences). Different portions of the overall anaglyph image were randomly selected and displayed on the screen. For each displayed area, observers viewed displays both before and after 3D refining. Before 3D refining, all observers had stereo fusion problems, i.e. impossibility and difficulty to fuse, for some areas. However, after 3D refining, depth perception was obvious over all areas, and observers did not have stereo fusion problems, even though *Y* parallaxes (maximum 2 pixels) occurred in some areas. Test results obtained in this work

indicate that the new 3D refining method significantly improve 3D effect display for the refined areas.

Refining time:

Refining time refers to the processing time requirement for the 3D refinement stage. In this research, a Quadtree-based method was proposed for 3D effect refining, and a modified bucket PR Quadtree (3DIndexTree) was implemented. The process of building the 3DIndexTree involves two steps: (1) loading the correspondence information from the file; and (2) building the 3DIndexTree. The building process is performed at the beginning of loading the image data, and so it does not really affect the refining time, but the search time on the 3DindexTree does. The refining time, therefore, includes both the search time and the time for refreshing the viewing area of the 3D image. The test result was 204 milliseconds for a 500 by 360 pixel viewing area; this result was the average of 20 separate runs for each trial. Because there is no requirement to transmit data over the Internet for 3D refining, and therefore no effects resulting from such transmission, it is asserted that testing using other machines online would have the similar results. This result also shows that efficient refinement of 3D the effect can be achieved, while meeting the real time requirement, using the proposed method.

4.2.3.2 Web Services of the Web-based 3D Visualization System

The performance of the architecture implementation of the W3VPS was tested, in terms of the efficiency of its functionalities, including zoom in, zoom out,

and pan. In the proposed approach, the 3DviewApplet applet cooperates with the Web server to provide those Web-based functionalities. This stage of the testing was repeated for a total of 20 trials without memory cleared between tests. The processing time for each test was measured from the instant that the client clicked a link to the Web server to the end of drawing an image (500 by 360 pixels) on the screen. The result obtained from this test, being the average processing time, was 748 milliseconds. This shows that the simple architecture for this research can provide good performance for those services.

4.3 Summary

The W3VPS developed in this research is completely automated. The matching, epipolar-generating, and anaglyph image-forming requires about 7 minutes to process in batch processing mode using a high-resolution satellite stereo image pair of around 4500 by 3000 pixels. The RMS of the matching results is 0.176 pixels in *xy*. The anaglyph image has a maximum of 2 pixels *Y* parallax and an average *Y* parallax RMS of 0.42 pixels. 3D images can be well displayed in the 3D viewers (see Figures 3-22 and 3-28 for the samples of display results) on both local computers and Web browsers. Functionalities such as 3D refining are achieved in real time. For Web visualization, data transmission over the Internet achieves good performance. It is like real time processing (by average time less then 748ms for a viewing area of 500 by 360 pixels).

Chapter 5

Conclusions and Recommendations for Future Work

In this research, the W3VPS has been designed, implemented, and tested. It allows: (1) automatical generation of 3D images through a new approach operating in anaglyph mode; and (2) 3D visualization of high-resolution satellite stereo images by means of Web browsers. The following sections summarize this thesis, highlight its chief contributions, and present some recommendations for future enhancements to the current W3VPS.

5.1 Conclusions

This thesis presents a new approach to the generation of natural colour 3D stereoscopic images using the anaglyph method based on human binocular matching sensitivity, and to the production of effective Web-based visualization. To better utilize the available high-resolution MS and Pan satellite images, and to explore a more effective way for detailed colour 3D visualization at low cost, the W3VPS requires pan-sharpened stereo images as data sources.

This study has demonstrated that it is possible to model a natural (urban) environment in colour 3D using the new commercial high-resolution satellite image data on the Web; the W3VPS allows the clear viewing of characteristically urban objects, including cars in colour 3D. The processing speed of colour 3D images such

as these is fast (around seven minutes for stereo images of 4000 by 3000 pixels), and colour 3D viewing can be accessed with the use of comparatively cheap complementary-colour glasses – an improvement over the expensive polarized screens and glasses required by traditional approaches to 3D viewing. The colour 3D image can be displayed on a conventional cathode-ray and active matrix screens and printed on paper - both characteristics that indicate great potential for a variety of applications and viewing environments.

Developing the W3VPS involved a number of innovative applications of photogrammetric principles to Web technologies. The contributions made by this thesis are summed up below.

1) A complete prototype system (W3VPS) was developed for automatic 3D image generation without the requirement of rigorous orientation parameters or DTMs; and for 3D Web visualization without special (expensive) graphics hardware that is typically required for high-resolution satellite stereo images. The W3VPS is easily accessible to the non-specialist. Through literature review and Web search, the author has found no similar systems in existence.

2) A new image-matching algorithm was developed for the automatic matching of corresponding pairs of high-resolution pan-sharpened stereo images. Image-tiling techniques were used to reduce the memory requirements, and to avoid having the established correspondences be too close to one another; image-tiling techniques also facilitate easy implementation of the algorithm and improve its performance. In this new image-matching algorithm, a modified SM method was

proposed and produced better results than those from Zhang's [1995] SM method. A new algorithm for eliminating outliers, based on the regression diagnostic approach, was developed and proven to be effective. This novel image-matching technique successfully and efficiently found a set of correspondences and eliminated the majority of reported correspondence outliers. Although not all outliers were rejected, the remaining few (1 unacceptable match over 142 good matches from the test pool) lay along their respective epipolar lines and had a negligible effect on evaluating epipolar image re-sampling.

3) A 3D image generation component was developed for the automatic generation of epipolar images as stereo image pairs, and for automatic formation of 3D anaglyph images. An algorithm was developed using a well-established 2D affine orientation model for the epipolar re-sampling of high-resolution satellite stereo images without the requirement to input rigorous orientation parameters, DTMs, or GCPs; this is a difference from the approach used in traditional digital 3D photogrammetric workstations and 3D satellite stereoscopic images systems. The algorithm's design entails no special prerequisite user knowledge and, thus, operation of the W3VPS is palpably simpler than existing counterparts. Based on epipolar images, high accuracy anaglyph images can be easily formed. The largest Y parallax of the resulting anaglyph images is less than 3 pixels.

4) A new algorithm based on the Quadtree technique was developed for refinement of the 3D effect. This new algorithm converts the 3D effect-refining problem to a range search in a Quadtree. By replacing traditional techniques (image-

matching), it improves the efficiency of refining the 3D effect online to meet the real time computing requirement. The search time averages 204 (ms).

5) GUIs were created to display 3D images on local computers or through Web browsers. These GUIs allow users to interactively view 3D and 2D colour satellite images. They also provide some specific functionality, such as a 3D effect-refining function, designed to enable users to view the optimal 3D effect in the viewing window, essentially in real time.

6) A one-tier Web client/server architecture was implemented for 3D Web visualization. Although a simple client/server architecture was employed, it was adequate for the application investigated; this simplicity enabled easy deployment of Web contents (files) to the Web server. To overcome the problem of bandwidth constraint and to improve system efficiency, the technology used in this research divides high-resolution images into bite-size tiles and delivers “pixels on demand”. The average data delivery time is less than 748 milliseconds for a 500 by 360 pixel viewing area.

5.2 Future Work

Although the goal of this thesis has been reached, there still exist a number of ways to improve the system performance and to extend the functionalities of the system for future development. This future work could include the following:

1) The precision of the estimation of the epipolar image re-sampling relies directly on the matched points. To improve the quality of estimation of the re-sampling, the accuracy of matched points should be increased. Using a subpixel-

precision corner detector can improve the accuracy of matched points. A false match is occasionally aligned with the epipolar line and will not be detected by the W3VPS because only the epipolar constraint is used for outlier rejection. This, of course, does not disturb the estimation of the epipolar image re-sampling. However, by exploiting other constraints, such as a disparity gradient limit, it is capable of detecting more outliers.

2) Data volume transmitted over the Internet is the key to improving system performance, to allowing users to view the images smoothly, and to reduce data transmission time. Integrating state-of-the-art image compression techniques, such as JPEG2000, into the W3VPS can contribute to reducing the data volume transmitted over the Web.

3) It is expected that the generation of 3D geometric models will allow for online extraction of 3D features. Geo-referencing of the 3D images is required for 3D feature extraction, and to integrate GIS data layers with the anaglyph image.

4) System functionality may also be expanded so as to enable the performance of geospatial data queries and, otherwise facilitating the delivery of a wider range of geospatial data services through Web browsers. Designing a Web-based multi-tier client/server model will achieve this.

References

Abel, D, T. Kerry, R. Ackland, and S. Hungerford (1998). "An exploration of GIS architecture for Internet environments." *Computer, Environment and Urban Systems,* Vol. 22, No. 1; pp 7-23.

Al-Rousan, N., P. Chen, G. Petrie, Th. Toutin, and M. J. Valadan Zoej (1997). "Automated DEM extraction and ortho-image generation from SPOT level-1B imagery." *Photogrammetric Engineering and Remote Sensing,* Vol. 63, No. 8, pp. 965-974.

Apache Software Foundation (2002). *Apache Server HTTP Project*, The Apache Software Foundation. [On-line], 01 November 2002, http://httpd.apache.org/

Asada, H., and M. Brady (1986). "The curvature primal sketch." *IEEE PAMI*, Vol. 8, pp. 2-14.

Baillard, C., and O. Dissard (2000). "A stereo matching algorithm for urban digital elevation models." *Photogrammetric Engineering & Remote Sensing,* Vol. 66, No. 9, pp. 1119-1128.

Barnett, V. and T. Lewis (1994). *Outliers in Statistical Data: 3rd edition*. John Wiley, New York.

Boomgaard, Rein van den, and Rik van der Weij (2001). "Gaussian convolutions, numerical approximations based on interpolation." *2001 Scale-Space and Morphology in Computer Vision, Third International Conference, Proceedings*, Vancouver, Canada, July 7-8.

Carr, Daniel B. (1993). "Production of stereoscopic displays for data analysis." *Statistical Computing and Statistical Graphics Newsletter*, April.

Chua, C., Y. Ho, and Y. Liang (2000). "Rejection of mismatched correspondences along the affine epipolar line." *Image and Vision Computing*, Vol.18, pp. 445-462.

Chung, K. and S. Tseng (2001). "New progressive image transmission based on quadtree and shading approach with resolution control." *Pattern Recognition Letters*, Vol. 22, pp. 1545-1555.

Cochran, S. D. and G. Medioni (1992). "3D surface description from binocular stereo." *IEEE Trans. PAMI*, Vol.14, No. 10, pp. 981-994.

Cook, Mark K., Bradley A. Peterson, and Gene Dial (2001). “IKONOS technical performance assessment.” *Proc. SPIE,* Vol. 4381, p. 94-108.

Do, K., Y. Kim, T. Uam, and Y. Ha (1998). “Iterative relaxational stereo matching based on adaptive support between disparities.” *Pattern Recognition,* Vol. 31, No. 8, pp. 1049-1059.

Doyle, S., M. Dodge, and A. Smith (1998). “ The potential of Web-based mapping and virtual reality technologies for modeling urban environments.” *Computer, Environment and Urban Systems,* Vol. 22, No. 2; pp 37-55.

El-Qawasmeh, Eyas (2003). “A quadtree-based representation technique for indexing and retrieval of image databases.” *J. Vis. Commun. Image R.* Vol. 14, pp. 340–357.

Fazekas, Attila, and Istvan Santa (2000). “Skeletonization on quadtree represented images”, *Acta Mathematica Academiae Paedagogicae Nyregyhaziensis,*Vol. 16, pp. 79-87.

Fritz, Lawrence W. (1996). “The era of commercial earth observation satellites.” *Photogrammetric Engineering and Remote Sensing,* Vol. 62, No. 1, pp. 39-45.

Forsyth, David A., and Jean Ponce (2003). *Computer Vision: A Modern Approach.* Pearson Education, Inc.

Fraternali, Piero (1999). “Tools and approaches for developing data-Intensive Web application: A survey.” *ACM Computing Surveys*, Vol. 31, No. 3, pp. 227-263.

Haala, Norbert, D. Stallmann, and C. Statter (1998). “On the use of multispectral and stereo data from airborne scanning systems for DTM generation and landuse classification.” *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, No. B4, pp.204-210

Horaud, R., F. Veillon, and T. Skordas (1990). “Finding geometric and relational structures in an image.” *Proceedings First European conference on Computer Vision*, pp 374-384, Antibes, France, April 1990

Kim, Taejung (2000). “ A study on the epipolarity of linear pushbroom Images.” *Photogrammetric Engineering & Remote Sensing,* Vol. 66, No. 8, pp. 961-966.

Klette, Reinhard, Karsten Schlüns, and Andreas Koschan (1998). *Computer Vision: Three-Dimensional Data from Images.* Springer-Verlag Singapore Pte. Ltd.

Jokinen, O. (1998). “Area-based matching for simultaneous registration of multiple 3-D profile maps.” *Computer Vision and Image Understanding*, Vol. 71, No, 3, pp. 431-447.

Landes, Steffen (1999). “ 3-D user interface for a Web-based information system access to virtual worlds via internet.” *GIM International*, Vol. 13, No. 1, pp. 76-79.

Li, R. (1998). “Potential of high-resolution satellite imagery for national mapping products.” *Photogrammetric Engineering and Remote Sensing*, Vol. 64, No. 12, pp. 1165-1169.

Liang, Tang, and C. Heipke (1996). “Automatic relative orientation of aerial images.” *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 1, pp. 47-55

Linden, Peter van der (2002). *Just Java 2, Fifth Edition*, Prentice Hall, U.S.A.

Manouvrier, M., M. Rukoz, and G. Jomier (2002). “Quadtree representations for storage and manipulation of clusters of images.” *Image and Vision Computing*, (article in press).

Medioni, G., and Y. Yasumuto (1986). “Corner detection and curve representation using cubic B-spline.” *IEEE Int. Conf, Robotics and Automation*, pp764-769

Morrison, K. W., and R. S. Purves (2002). “Customizable landscape visualizations implementation, application and testing of a Web-based tool.” Computers, Environment and Urban Systems, Vol. 26, pp 163-183.

Mount, D. M., N. S. Netanyahu, and J. L. Moigne (1999). “Efficient algorithm for robust feature matching.” *Pattern Recognition*, Vol. 32, pp. 17-38.

Okamoto, Atsushi, C. Fraser, H. Hattori, and T. Ono (1998). “An alternative approach to the triangulation of Spot imagery.” *IAPRS*, Vol. 32, *Part 4 “GIS-Between Visions and Applications”, Stuttgart.*

O’Neil, M. A. and I. J. Dowman (1988). “The generation of epipolar synthetic stereo mates for SPOT images using a DEM.” *International Archives of Photogrammetry and Remote Sensing*, Vol. 27, No. B3, pp.587-598.

O’Neill, M. and M. Denos (1996). “Automated system for coarse-to-fine pyramidal area correlation stereo matching.” *Image and Vision Computing*, Vol. 14, No. 3, pp.225-236

Ono, Tetsu (1999). “Epipolar resampling of high resolution satellite imagery.” *Proceedings, ISPRS Workshop "Sensors and Mapping form Space 1999"*, Nr. 18, Institut für Photogrammetrie und Ingenieurvermessungen, Universität Hannover September 27-30.

Orfali, Robert, Dan Harkey, and Jeri Edwards (1996). *The Essential Client/Server Survival Guide: Second Edition.* John Wiley & Sons, Inc. New York.

Otto, G. P. (1988). “Rectification of SPOT data for stereo image-matching.” *International Archives of Photogrammetry and Remote Sensing*, Vol.27, No. B3, pp.635-645.

Patterson, Robert (1992). “Human stereopsis.” *Human Factors*, Vol. 34, No. 6, pp. 669-692.

Patterson, Robert (1997). “Visual processing of depth information in stereoscopic displays.” *Displays*, Vol. 17, pp 69-74.

Petrie, G. (1999). “Automatic DEMs and orthoimages from SPOT stereo-pairs: SPOT DEMs of Jordan.” *GIM International*, Vol, 13, No. 10, pp. 76-79.

Prade, La G., S. J. Briggs, R. J. Farrell, and E. S. Leonardo (1966). *Stereoscopy*. in Manual of Photogrammetry (Chapter X), Third Edition, (Bethesda, USA: ASPRS), pp. 519-544.

Samet, Hanam (1990). *The Design and Analysis of Spatial Data Structure.* Addison-Wesley Publishing Company, Inc.

Sadoski, Darleen (2000). *Client/Server Software Architectures - An Overview*. The Software Engineering Institute at Carnegie-Mellon University [On-line], 22 September 2000, http://www.sei.cmu.edu/str/descriptions/clientserver_body.html

Samet, H. and R. E. Webber (1988). “Hierarchical data structures and algorithms for computer graphics. Part I: fundamental.” *IEEE Computer Graphics and Applications*, Vol. 8, No. 5, pp. 48-68.

Schenk, Toni (1999). *Digital Photogrammetry*. TerraScience, Laurelville.

Sebesta, Robert W. (2002). *Programming the World Wide Web.* Addison Wesley Longman.

Shaffer, C. A., H. Samet, and R. C. Nelson (1990). “QUILT: a geographic information system based on quadtree.” *International Journal on Geographical Information Systems*, Vol. 4, No. 2, pp. 103-131.

Shapiro, L. S., and J. M. Brady (1995). "Rejecting outliers and estimating errors in an orthogonal-regression framework," *Phil Trans Roy Soc Lond*, Vol. 350, pp. 407-439.

Stevenson, S. B., and C. M. Schor (1997). “Human stereo matching is not restricted to epipolar line.” *Vision Res.*, Vol. 37, No. 19, pp. 2717-2723.

Sugumaran, R., C. Davis, J. Meyer, T. Prato, and C. Fulcher (2000). “Web-Based decision support tool for floodplain management using high-resolution DEM.” *Photogrammetric Engineering & Remote Sensing*, Vol. 66, No. 10, pp 1261-1265.

Sun Microsystems, Inc. (2002). *Code Samples and Apps: Applets*. Sun Microsystems, Inc. [On-line], 18 November 2002. http://java.sun.com/applets/

Sun Microsystems, Inc. (2003a). *Java™ Advanced Imaging (JAI) API*. [On-line], 24 February 2003, http://java.sun.com/products/java-media/jai/

Sun Microsystems Inc. (2003b). *JDK1.4 Programming Environment*. [On-line], 24 November 2003, http://www.sun.com

Toutin, T. (2001). “Review article: elevation modeling from satellite visible and infrared (VIR) data.” *International Journal of Remote Sensing*, Vol. 22, No.6, pp. 1097-1125.

Toutin T., and Cheng P. (2002). “QuickBird - A milestone for high resolution mapping.” *EOM*, Vol. 11, No. 4, pp. 14-18.

Toutin T., R. Chénier, and Y. Carbonneau (2002a). “3D models for high resolution images: Examples with Quickbird, IKONOS, and EROS.” *ISPRS Commission IV Symposium, Joint International Symposium on Geospatial Theory, Processing and Applications,* Ottawa, July 8-12, Vol. 34, No 4, pp. 547-551.

Toutin T., R. Chénier, Y. Carbonneau, and N. Alcaïde (2002b). “3D mapping with high resolution images.” *22nd EARSeL Symposium, Prague, Czech Republic*, June 4-6.

Toutin, T. and Corinna Vester (2003). *Remote Sensing Tutorials: Radar and Stereoscopy*, Canada Centre for Remote Sensing. [On-line], 12 November 2003, http://www.ccrs.nrcan.gc.ca/ccrs/learn/learn_e.html

Wang, Ching Yang, Shih Jac Liao, and Long Wen Chang (2000). "Wavelet image coding using variable blocksize vector quantization with optimal quadtree segmentation." *Signal Processing: Image Communication,* Vol. 15, pp. 879-890.

Wang, Shaowen and Marc P. Armstrong (2003). "A quadtree approach to domain decomposition for spatial interpolation in grid computing environments." *Parallel Computing*, Vol. 29, pp. 1481–1504.

Weng, J., N. Ahuja and T. S. Huang (1992). "Matching two perspective views." *IEEE Trans. PAMI*, Vol. 14, No. 8, pp. 806-825.

Wolf, Paul R. and Bon A. Dewitt (2000). *Elements of Photogrammetry with Applications in GIS: 3rd Edition.* McGraw-Hill, Boston.

Yang, Yuh-Horng, Kuo-Liang Chung, and Yao-Hong Tsai (2000). A compact improved quadtree representation with image manipulations." *Image and Vision Computing,* Vol. 18, pp. 223–231.

Zhang, Y. (2002a). "Natural colour urban 3D modeling: A stereoscopic approach with IKONOS multispectrsal and panchromatic data." *International Archives of Photogrammetry and Remote Sensing (IAPRS),* Volume 34, Part 4 (ISPRS "GeoSpatial Theory, Processing and Applications", Ottawa, July 2002).

Zhang, Y. (2002b). "A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images." *Proceedings of IEEE/IGARSS'02*, Toronto, Canada, June 24-28.

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

Zhou, J., and J. Shi (2002). "A robust algorithm for feature point matching." *Computers and Graphics,* Vol. 26, pp. 429-436.

Vita

Candidate's full name: Pingping Xie

University attended: B.Eg (Cartography), 1990
Wuhan Technical University of Surveying and Mapping
Hubei, P.R. China