TEXTURE MAPPING OF BUILDING FAÇADES USING OBLIQUE TERRESTRIAL PHOTOGRAPHY

ALLAN TAN YONG KIAT

NATIONAL UNIVERSITY OF SINGAPORE

2009

TEXTURE MAPPING OF BUILDING FAÇADES USING OBLIQUE TERRESTRIAL PHOTOGRAPHY

ALLAN TAN YONG KIAT
(B.Eng.(Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

I would like to thank my supervisor, Associate Professor Ong Sim Heng for his guidance on the research work done in this thesis. His enthusiasm is a great driving force which ensured the smooth progression of my work.

My sincere gratitude to Mr Kwoh Leong Keong for his ideas and direction which helped to evolve the project to its final stage. His feedback and comments during the seminar series held were most appreciated.

I would like to thank the following people from the Centre for Remote Imaging, Sensing and Processing (CRISP) who had rendered help and support in course of this project: Ms Huang Xiaojing for her RPC code; Mr Mak and his ground station crew for the satellite imagery and 3-D building data; Mr Tan Wee Juan for programming advice. Without which this project would not have been able to achieve its objectives.

I would like to thank Mom and Dad for their constant support and encouragement throughout the entire journey. There are no words to describe my indebtedness to you folks.

Finally, I would like to thank Lisa, my pillar of support whose kind understanding and patience helped me through all the problems encountered. Thank you. Thank you. Thank you.

Table of Contents

Summary

The focus of this work is to create photorealistic façade textures using oblique view photography for building models in a virtual city model environment. Building models are extracted using remote sensing means and we have accurate data on the building shape and height. On the other hand, building façade details are lacking due to sensor limitations, hence we can improve the amount of details by texturing the façade with terrestrial photographs. An Improved Reverse Projection Texture Mapping (IRPTM) core was implemented in a computer program named Graffiti to process the terrestrial photographs and extract façade textures. IRPTM involves the recovery of the image acquisition parameters, the automatic extraction of façade textures, the geometric correction of distorted façade textures and the colour balancing of the final façade texture composition. In this thesis, experimental results will show that IRPTM is a better technique as compared to both current Forward Texture Mapping (FTM) methods and Reverse Projection Texture Mapping (RPTM) methods to perform texturing of building models in a quick and efficient manner.

The oblique view terrestrial photographs are taken without the help of external direct referencing sensors and we require the camera's extrinsic parameters in order to extract the textures automatically. We have developed two methods to recover the extrinsic parameters belonging to the camera at the time of acquisition, namely Hybrid Space Resection method and Interactive Pose Estimation method. Hybrid Space Resection combines two existing methods, closed-form space resection and iterative non-linear space resection to recover the camera extrinsic parameters using 4 sets of control points. By varying the location and orientation of the OpenGL camera, Interactive Pose Estimation uses the reverse projection of façade textures onto the terrestrial photograph to aid the user in the recovery of parameters. Both methods were

shown to accurately recover the extrinsic parameters of the camera in our experiments conducted.

Oblique photography of building façades introduces a skew distortion to the extracted textures. This is due to the inherent perspective nature of the camera used for acquisition. Geometric correction via OpenGL re-projection onto a quadrilateral can be carried out and was shown to preserve the geometric accuracy of the façade texture. Radiometric differences in the terrestrial photographs are introduced to the extracted textures. This is due to the different illumination conditions and imaging equipment used in the acquisition stage. It was shown that we can improve the colour continuity between adjacent façade textures by performing colour balancing.

List of Tables

List of Figures

List of Symbols

φ	Angle of rotation about the X axis
ω	Angle of rotation about the Y axis
κ	Angle of rotation about the Z axis
X, Y, Z	3-D object coordinates in real world reference frame
x, y	2-D image coordinates in base image reference frame
x', y'	2-D image coordinates in non-base image reference frame
f	Focal length
m	Rotation matrix (3 x 3)
FOV_x	Horizontal field of view angle
FOV_y	Vertical field of view angle
d	Imaging sensor height in mm
mm	Millimeters
m	Meters
rad	Radians
width	Width of OpenGL rendered scene on screen
height	Height of OpenGL rendered scene on screen
U, V	2-D texture coordinates
p	8 bit pixel intensity
μ	Mean intensity
σ	Standard deviation of intensity
h	3x3 plane projection transformation matrix

Abbreviations

3-D	3 Dimensional
2-D	2 Dimensional
CMOS	Complementary Metal Oxide Semiconductor
MR	Mixed Reality
LIDAR	Light Detection and Ranging
RPC	Rational Polynomial Coefficient
FTM	Forward Texture Mapping
RPTM	Reverse Projection Texture Mapping
IRPTM	Improved Reverse Projection Texture Mapping
AOI	Area Of Interest
GPS	Global Positioning System

Chapter 1

Introduction

The work presented in this thesis focuses on improving the photo-realistic quality of building models in a virtual 3-D city environment. In the course of this chapter, the brief concepts of photogrammetry and mixed reality will be introduced. This will be followed by a review of the potential texture mapping applications for building models, making apparent the increasing interest in 3-D city model visualization platforms. We end this chapter by presenting an overview of this thesis and its organization.

1.1 Photogrammetry

Photogrammetry is the science of obtaining reliable information about physical objects and their surrounding environment through the processes of recording, measuring and interpreting photographic images. Historically, photogrammetry progressed from black and white terrestrial photographs taken on light sensitized silver iodide coated metal plates to ultra high resolution, 16-bit colour digital images taken from advanced Complementary Metal Oxide Semiconductor (CMOS) sensors mounted on space-borne satellites. As implied, this is the science of analyzing film and digital photographs, radiated acoustic energy patterns as well as magnetic inferences. The process of making precise measurements from photographs and other information sources to determine relative locations of points is known as metric photogrammetry.

This aids in the finding of distances, angles, areas, volumes, elevations and sizes and shapes of objects.

There are two fundamental types of photographs used in the photogrammetry, namely terrestrial and aerial as shown in Figures 1.1 and 1.2, respectively. Terrestrial photographs are essentially images taken using ground based cameras, from which the position and orientation of the cameras are often measured directly at the time of image acquisition. These terrestrial photographs can be taken from a variety of cameras ranging from simple hand-held point-and-shoot cameras, single lens reflex cameras to precise surveying cameras mounted on tripods. Aerial photographs on the other hand are images taken from large and expensive cameras mounted on airborne or space-borne platforms. These are equipped with highly accurate sensors to provide precise measurements of the position and orientation at the time of acquisition.

Figure 1.1: Example of a terrestrial photo.

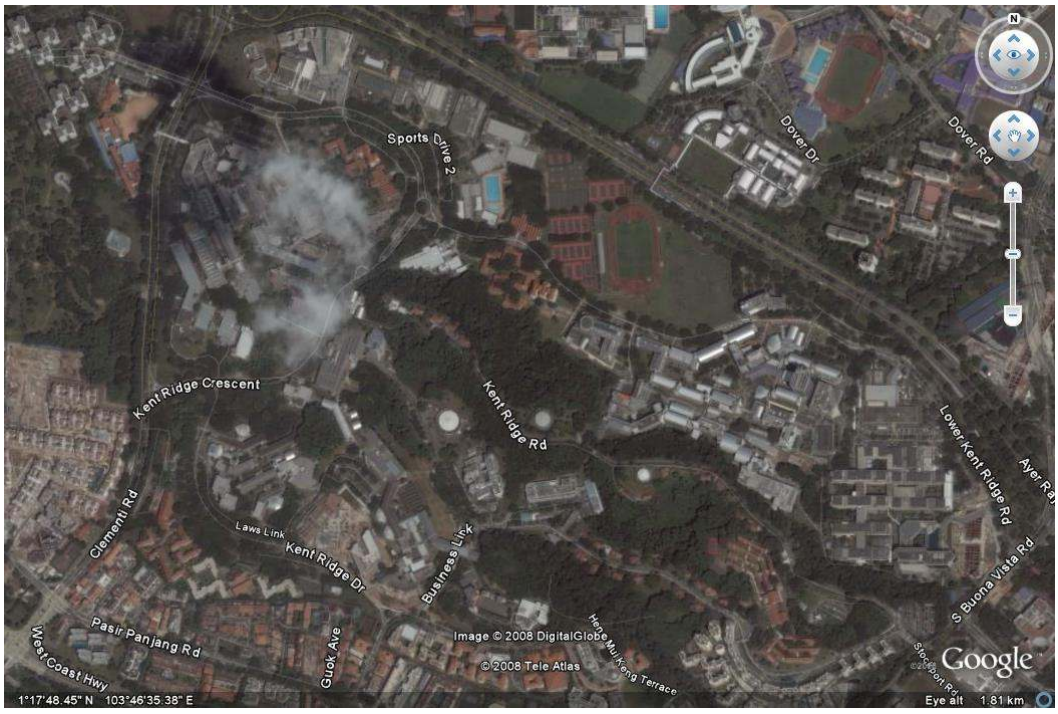

Figure 1.2: Example of a satellite image taken by QUICKBIRD satellite.

1.2 Mixed Reality

Mixed Reality (MR) as defined by Milgram and Kishino (1994) is the region of continuum between the completely real world through to the completely virtual environment as shown in Figure 1.3. The conventional view of the mixed reality environment lets the user be partially or totally immersed in, and able to interact with a completely synthetic world which may mimic the properties of some real-world environments, either existing previously or currently. The most straight forward way to understand a Mixed Reality (MR) environment would be one which allows real and virtual objects to be presented together within a single display. By augmenting a virtual environment with real objects, we are able to provide an integrating medium to merge richly layered, multi-modal, 3-D real context and experience together.

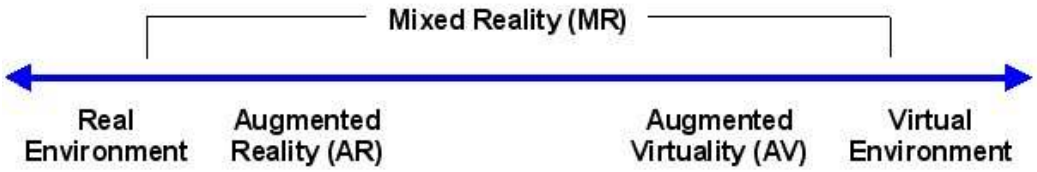

Figure 1.3: Reality-Virtuality continuum.

By using MR systems, users would be able to better visualize their living spaces with greater detail through in-depth interactions as virtual objects resemble their real world counterparts. They can be manipulated in a virtual setting without being constrained by logical, spatial and temporal limitations, allowing human activities to be modeled more intuitively and efficiently than before. A virtual 3-D city model would be an ideal application of MR platforms as it is a sophisticated application of geo-informatics systems, representing many layouts, activities and functionalities of a real-world community. It is an integrated effort in the fields of computer graphics, remote sensing and engineering to model the appearance and dynamics of the real world. The primary aim would be to introduce a feasible, yet compelling representation of the city in the virtual world. It will be an attractive option to view and explore the building architecture of the past, present and future without being encumbered by the constraints of reality.

1.3 Photogrammetry and Mixed Reality

In the context of a MR city visualization platform, accurate reconstruction and realistic texture mapping of both 3-D terrain and building models surfaces would enhance the immersive experience of the user. Baltsavias *et al.* (2001) provides a good overview for the collection of 3D extraction techniques ranging from high resolution aerial

stereo images to airborne laser scanners (LIDAR) using various semi-automated and automated techniques as shown in Figures 1.4 and 1.5.

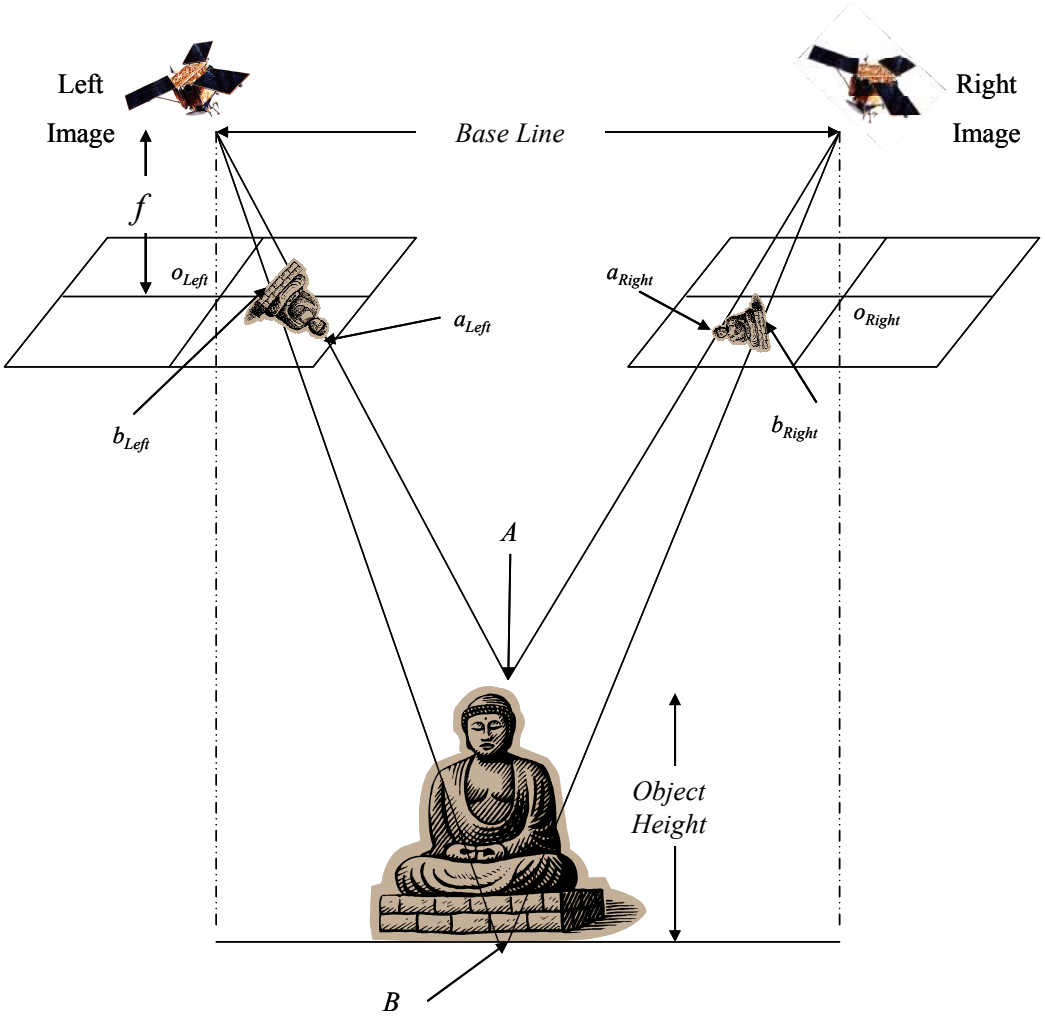

Figure 1.4: Stereo pair imaging by satellite.

In the stereo setup, the height of the ground object as shown in Figure 1.4 can be recovered using Eqs. (1.1) and (1.2) (Mikhail *et al.*, 2001):

$$Object\ Height = Base\ Line \times f \times \left[\frac{P_{Right} - P_{Left}}{P_{Left} P_{Right}} \right] \tag{1.1}$$

$$\begin{aligned} P_{Left} &= a_{Left} + a_{Right} \\ P_{Right} &= b_{Left} + b_{Right} \end{aligned} \tag{1.2}$$

where a_{Left} are the image coordinates of pt A in the Left image

b_{Left} are the image coordinates of pt B in the Left image

a_{right} are the image coordinates of pt A in the Right image

b_{Right} are the image coordinates of pt B in the Right image

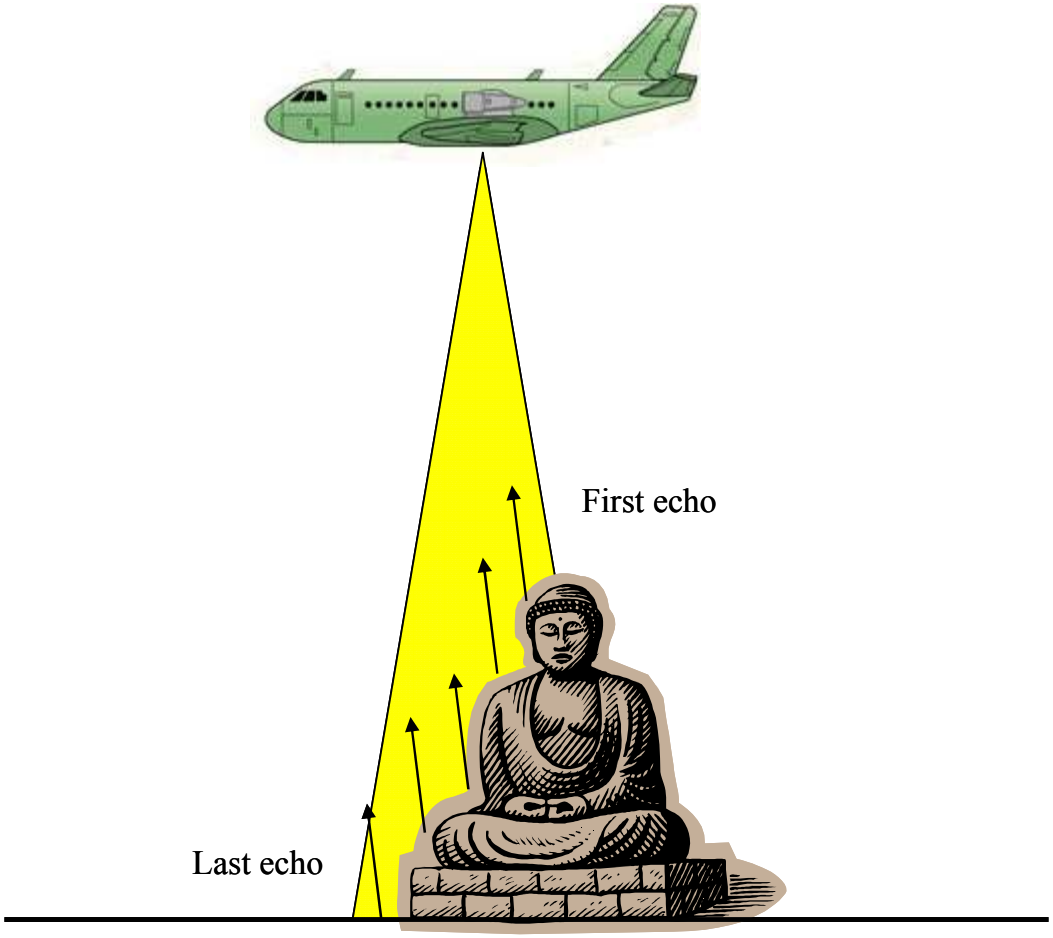

Figure 1.5: Elevation data collection by airborne LIDAR.

The Rational Polynomial Coefficient (RPC) camera model is a mapping function which uses ratios of cubic polynomials to express the transformation from ground surface coordinates (latitude, longitude and elevation) to image coordinates (line, pixel) for the satellite image. Using current high resolution satellite imagery coupled with RPC camera model, we are able to accurately estimate 3-dimensional building information (Grodecki, 2001). In the case of stereo images from the IKONOS

satellite, the accuracies of the extracted features are within 1 meter horizontally and 2 meters vertically. This is acceptable for use in urban mapping whereby building objects' dimensions are usually bigger than 3 meters (Grodecki and Dial, 2001).

Airborne and satellite remote sensing data can provide accurate 3-D information on urban building objects but limited texture information. High resolution LIDAR is unable to provide any useful texture information as the data collected constitutes a dense pattern of points which represent the return intensities of the laser pulses. Furthermore, the amount of texture information which we can extract from high resolution optical satellite imagery is limited to building tops and certain building façades as there is a restriction of sensor look angles during image acquisition. The extracted building façades are pixelated due to the lack of resolution and texture information (Huang *et al.*, 2006) as the building façade which can be seen from the satellite imagery is relatively small when compared to the actual size. Therefore, it is observed that current 3D building models extracted from remote sensing means do not have sufficient textural information about the building façades.

With the release of virtual globe programs which allow users to view satellite maps and interact with 3-D city models, there is an increasing public interest in geospatial technologies and applications like Google Maps, MapQuest, GoogleEarth and other 2-D and 3-D visualization platforms (Scharl, 2007). These MR city model visualization platforms often use OpenGL or DirectX to effectively render in real-time buildings models and ground terrain using hardware acceleration. When texture information of the façade belonging to a building model is not available, it is a common practice to assign a pseudo texture instead (Beck, 2003) as shown in Figure 1.6. Although this provides an adequate presentation of the building model, it is not ideal as it does not portray the true characteristics of the building façade. Such pseudo

representations of the building facades make the overall visualization system less realistic and less useful. In applications like automated pose estimation of terrestrial photographs taken in an urban environment, the lack of façade textures on the buildings would greatly reduce the effectiveness of using such a visualization system to generate images for possible pose matching.

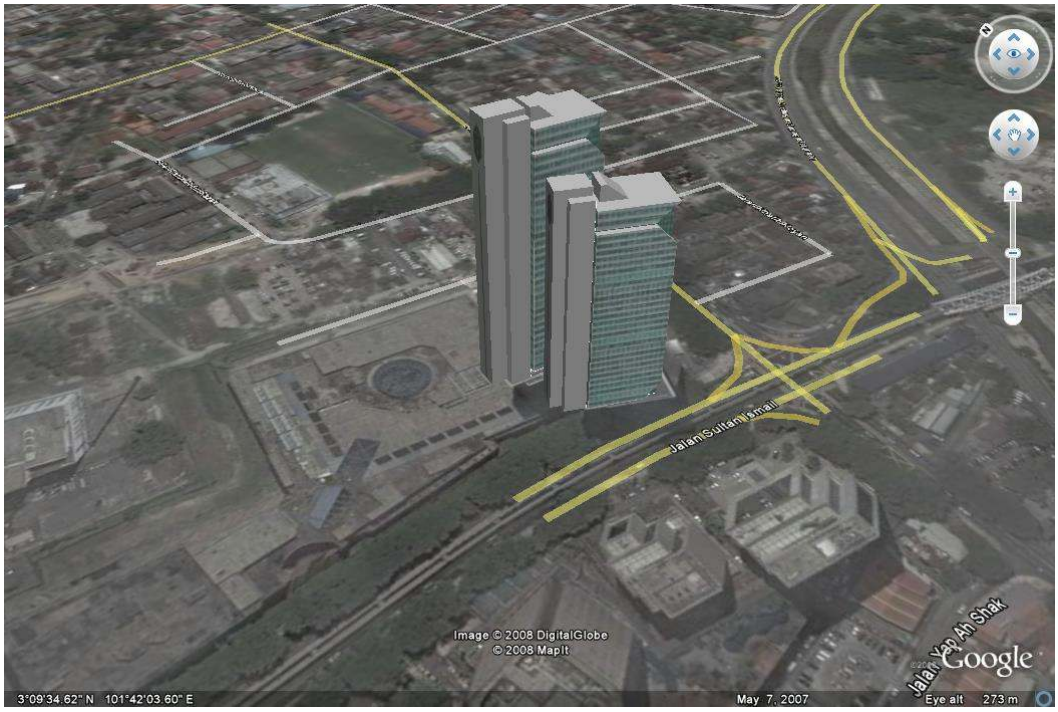

Figure 1.6: Example of a building model textured with a pseudo blank texture.

There are several building model texturing programs currently available to the users in the market. They include community supported software like Sketchup and Blender, and more commercially-driven specific applications like Autodesk 3ds Max and Revit Architecture. The cost of software can range from nothing to a couple of thousand dollars due to the additional features incorporated into the software. They are often used in the fields of computer animation, game design and building design.

1.4 Overview

The aim of this work is to study and generate photo-realistic façade textures for the building models to improve the visual quality and usability of the visualization platform. These textures can be generated from terrestrial digital photographs taken using commercially available cameras. The process of allocating and mapping the appropriate texture from terrestrial photographs to correct building façade is both laborious and time-consuming. To address this issue, we have developed a system called Graffiti which implements a semi-automated process for the large scale texturing of the building façades belonging 3-D building models previously extracted from remote sensing techniques.

1.5 Thesis Organization

The organization of this thesis is as follows:

In Chapter 1, a brief introduction of photogrammetry as well as mixed reality is given. This is followed by a section describing how photogrammetry can aid in the effective reproduction of building models in a MR visualization platform and its applications.

A brief review on the current methods used for texture mapping is presented in Chapter 2, namely Forward Texture Mapping (FTM) and Reverse Projection Texture Mapping (RPTM). This gives the reader an insight of the various working principles adopted previously by other approaches.

The problem statement of which this thesis will be answering is described in Chapter 3, followed by the limitations of the current methods employed for texture mapping of building models. The reasons which support the validity for the development of an improved method to perform texture mapping is elaborated as well.

The methods used in this system are described in greater detail in Chapter 4. The methods are:

- Texture mapping
- Determination of camera extrinsic parameters
- Geometric correction of façade textures
- Colour balancing of façade textures
- Mosaic composition of multiple images

Chapter 5 evaluates the performance of our improved texture mapping technique. Real terrestrial images are taken and used to texture building models which were previously obtained via remote sensing means. In addition, a few data sets comprising control points and camera orientation and location will be used to evaluate the accuracy of the camera locating algorithm.

Finally, Chapter 6 summarizes the achievements of this work and concludes with suggestions for future work to improve the overall schema of texture mapping building models for visualization purposes.

Chapter 2

Review of Current Methods

In the context of 3-D city model visualization, the technique of mapping grey-scale or true colour images onto the façade surfaces belonging to a 3-D building model is known as texture mapping. This is done in order to improve the photo-realistic quality of the augmented object in a MR environment. The integration of multiple digital images is necessary for the complete generation of the façade textures belonging to the building model. There are many approaches presented in literature which perform the integration of texture mapping operations and can be summarized to two broad methods. These methods are namely:

1. Forward Texture Mapping (FTM)
2. Reverse Projection Texture Mapping (RPTM)

2.1 Forward Texture Mapping (FTM)

In general, Forward Texture Mapping (FTM) method is a mapping transform which registers 2-D texture space to 3-D model space without the need for the camera's acquisition location and orientation. Area of interest (AOI) outlines of the facades must be allocated as polygons on the photographs either manually or automatically (Zhang and Kang, 2004; Laycock and Day, 2004) before these textures can be mapped onto the corresponding building model facades to produce a complete and accurate representation of the building as shown in Figure 2.1.

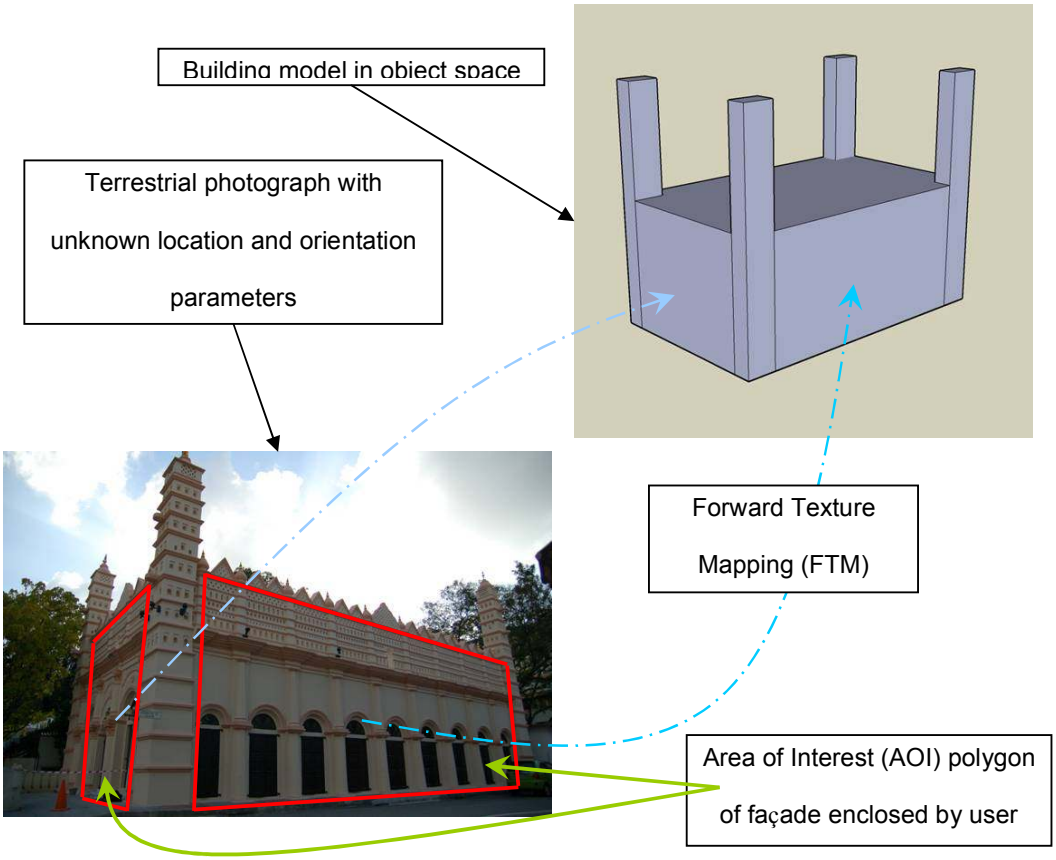

Figure 2.1: Example of Forward Texture Mapping (FTM).

Some approaches in literature make use of multiple images from a static location or dynamic moving camera to create a seamless and geometrically continuous mosaic (Coorg and Teller, 2000; Nicolas, 2001; Zhang and Kang, 2004; Tsai and Lin, 2007) of the building façades to achieve realistic and impressive results. Other methods make use of edge detection algorithms for the automatic detection and generation of repeatable wall patterns for the texturing of the building façade (Laycock and Day, 2004).

Coorg and Teller (2000) demonstrated that spherical mosaics can be generated from a series of images acquired from a common optical center. The camera acquires images at pre-determined locations using external sensors for the initial estimation of orientation. On the average, this technique uses about 50 images per location to produce a spherical mosaic of acceptable quality. By using a spherical mosaic optimization technique, an image with an effectively super hemispherical field of view can be produced, eliminating ambiguity between camera translation and rotation as shown in Figure 2.2. The resultant mosaic can be treated as a rigid composite image and projected onto a cubical environment, preserving the straight edges of the source images and allowing the building façade textures to be extracted.

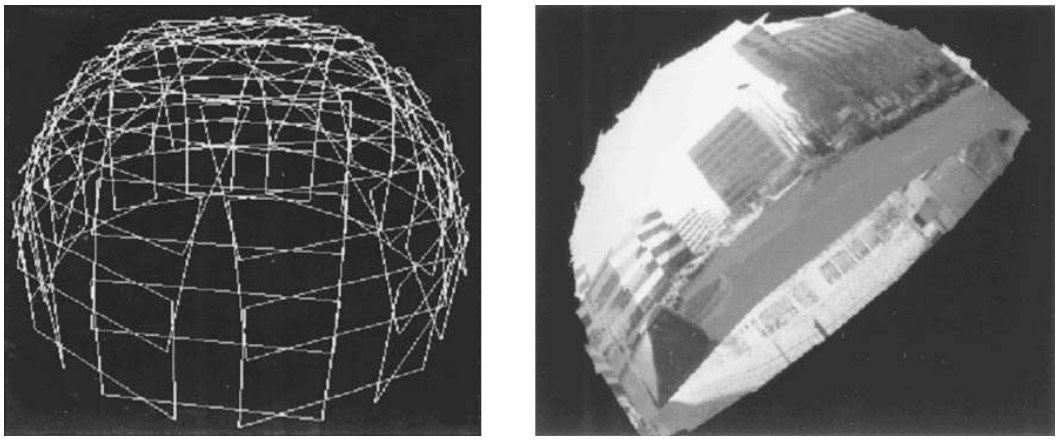

Figure 2.2: Super hemispherical field of view for a node of the dataset (Coorg and Teller, 2000).

Instead the static acquisition of images, Nicolas (2001) showed that dynamic mosaic representations of a scene can be generated using object-based manipulations of the data obtained from a moving video camera. With a combination of affine transformation based on warping parameters estimation and an analytic criterion for image blending, high quality mosaic sequences can be created. This can be done for off-line applications once the entire video sequence is available for processing.

Zhang and Kang (2004) described an automatic mosaicking process from land-based video sequences. Registration control point pairs were automatically selected by correlation from the sequence of video frames. Bundle adjustment of the image was carried out after combining the constraints of straight lines bundle and known orientation of parallel lines in object space. Bundle adjustment is the process which adjusts all image measurements to ground control values in a single solution. The process is named because of the many light rays that pass through each lens position constituting a bundle of rays. The bundles from all photos are adjusted simultaneously so that corresponding light rays intersect at positions of the control points. The adjusted image frames were subsequently merged together, forming a mosaic composition. Furthermore, by assuming that object lines belonging to building façades were parallel to the vertical and horizontal axis in object space, Zhang and Kang were able to combine this criterion with the image's hue information and effectively remove parts of the image occluded by vegetation as shown in Figure 2.3.

Figure 2.3: Vegetation occlusion removal by using hue information (Zhang and Kang, 2004).

The process of performing polygon AOI outlining of the façade on the mosaic composition is a simple but labour intensive task. Tsai and Lin (2007) tried to improve the workflow of the entire process by introducing a polygon-based technique for the texture mapping of building models. Textures of building exteriors were generated from mosaics of close-range photographs acquired from commodity digital cameras. Multiple digital photographs were merged together using minimum 8 manually selected control points for the registration of the overlapping regions of the photographs. Using least-squares fitting, the images were transformed from individual image spaces to a common texture space, forming the mosaic composition as shown in Figures 2.4 and 2.5. The building textures were then extracted from the mosaic using polygonization of the façade outline, a process of creating a polygon AOI outline on the image and assigning it to the respective building façade.

Figure 2.4: Three texture images to mosaic (Tsai and Lin, 2006).

Figure 2.5: Resultant mosaic of continuous façade after using control points for merging (Tsai and Lin, 2006).

Laycock and Day (2004) proceed to perform the polygonization of the façade outline automatically with a certain degree of success. They showed that by using a shrinking polygon which was based on the principle of active contour models in machine vision (Kass *et al*., 1987), the façade object on the texture can be extracted automatically. Prior image processing using the Canny edge detection algorithm (Canny, 1986) was carried out on the digital photograph to create a façade edge map. The polygon was shrunk in a series of iterative cycles so that it will come into contact with the extracted façade edges. By performing line intersection test between the polygon edges and the façade edge map, the polygon will cease to shrink upon contact on further iterative cycles. After joining all the stopped edges, the façade polygon outline was extracted automatically.

2.2 Reverse Projection Texture Mapping (RPTM)

The Reverse Projection Texture Mapping (RPTM) method is similar to the Forward Texture Mapping (FTM) in principle. It is a mapping transform which registers 2-D texture space to 3-D model space. The core difference between both methods is in the AOI texture extraction process of the façade from the image. RPTM method requires all the intrinsic and extrinsic parameters belonging to the camera which was used for the image acquisition process. As shown in Figure 2.6, intrinsic parameters are the inherent characteristics of the digital camera and usually consist of focal length, pixel resolution, principal center, image dimensions and lens distortions. Extrinsic parameters of the camera comprise of the location and orientation of the camera with reference to the world reference frame at the instant of image acquisition.

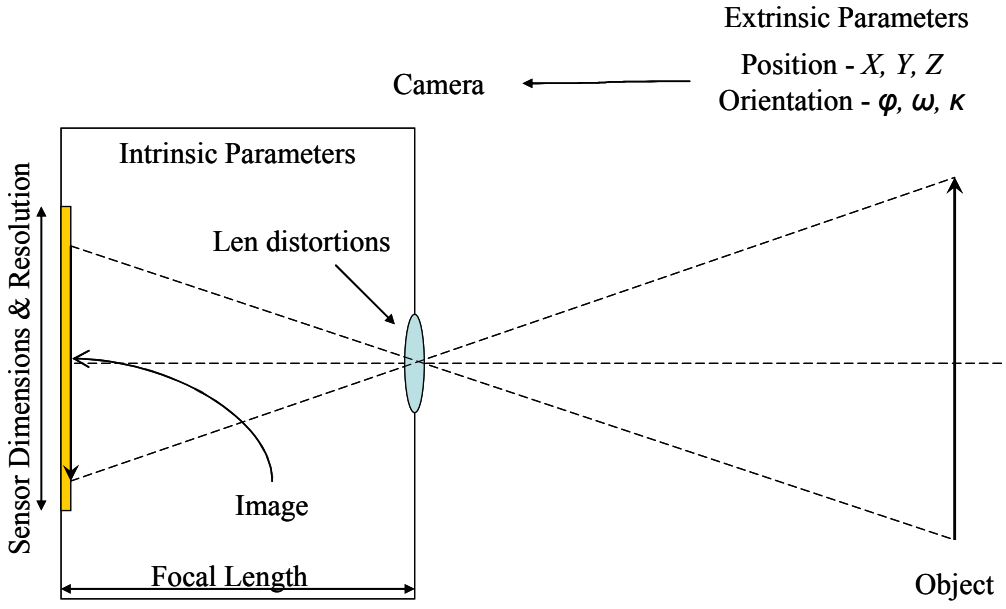

Figure 2.6: Extrinsic and intrinsic parameters of a digital pinhole camera model.

The relationship between the image acquisition station, any object point and its corresponding mapped point on the image can be seen in Figure 2.7 and described by the following equations (McGlone *et al*., 2004):

$$x_a = x_0 - f\left[\frac{m_{11}(X_A - X_C) + m_{12}(Y_A - Y_C) + m_{13}(Z_A - Z_C)}{m_{31}(X_A - X_C) + m_{32}(Y_A - Y_C) + m_{33}(Z_A - Z_C)}\right] \tag{2.1}$$

$$y_a = y_0 - f\left[\frac{m_{21}(X_A - X_C) + m_{22}(Y_A - Y_C) + m_{23}(Z_A - Z_C)}{m_{31}(X_A - X_C) + m_{32}(Y_A - Y_C) + m_{33}(Z_A - Z_C)}\right] \tag{2.2}$$

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \begin{bmatrix} \cos\varphi\cos\kappa & \sin\omega\sin\varphi\cos\kappa & -\cos\omega\sin\varphi\cos\kappa + \sin\omega\sin\kappa \\ -\cos\varphi\sin\kappa & -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa & \cos\omega\sin\varphi\sin\kappa + \sin\omega\cos\kappa \\ \sin\varphi & -\sin\omega\cos\varphi & \cos\omega\cos\varphi \end{bmatrix} \tag{2.3}$$

where x_a, y_a are the image coordinates of the image pt a
X_A, Y_A, Z_A are the object space coordinates of A
X_C, Y_C, Z_C are the object space coordinates of station C
$m_{11} - m_{33}$ are the components of the orientation matrix
f is the focal length
x_0, y_0 are the image coordinates of the pixel centre
ϕ, ω, κ are the rotations about the X, Y and Z axis respectively

The intrinsic and extrinsic parameters are essential in the application of the collinearity equations as they would affect the position of the transformed image point.

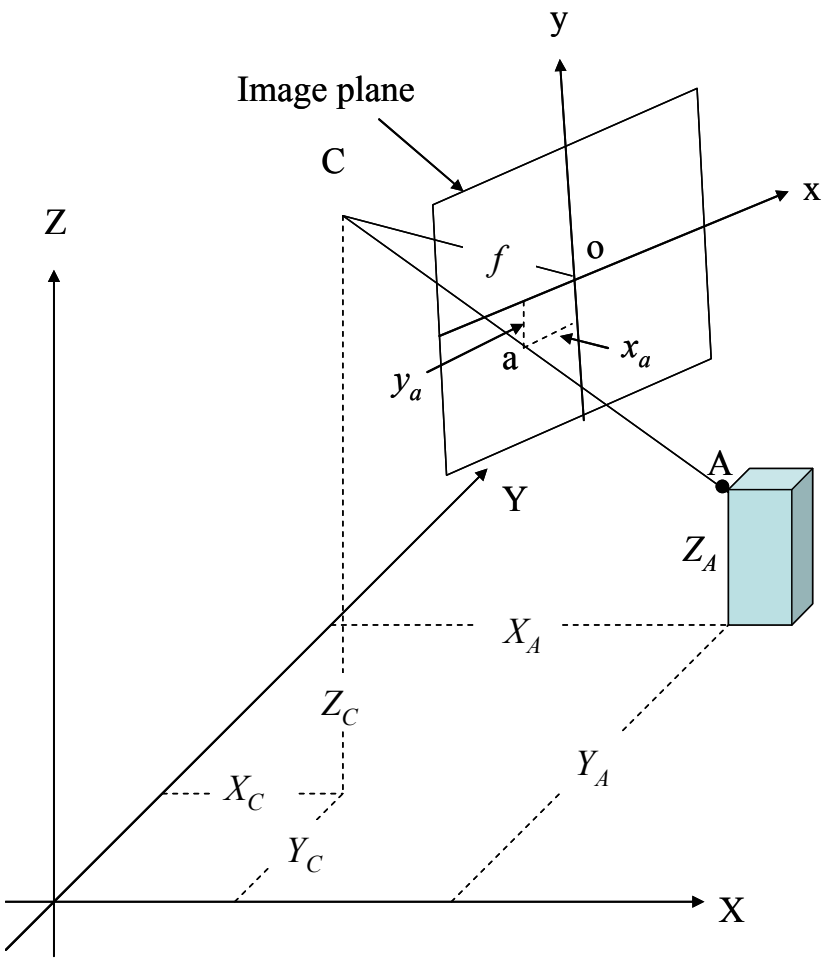

Figure 2.7: Visual interpretation of collinearity equations 2.1 and 2.2 in 3-D space.

The recovery of the camera's location and orientation is vital for the accurate extraction of the façade textures for RPTM. As the complete texturing of a typical building often requires multiple images taken from different view points, several methods have been developed to either impose limits (Coorg and Teller, 1999) or estimate (Debevec *et al*, 1996; Kumar *et al*., 2000; Coors *et al*, 2000; Haala and Bohm, 2003) the extrinsic parameters of the camera via direct or indirect means.

Coorg and Teller (1999) showed that by taking 4000 high resolution, close-range images of buildings from fully controlled positional nodes with known extrinsic parameter information, they were able to fully automate the vertical texture extraction

process for a small area of 200 square meters. This was based on the assumption that the imaging plane differs from the façade plane by a single azimuth rotation as shown in Figure 2.8.

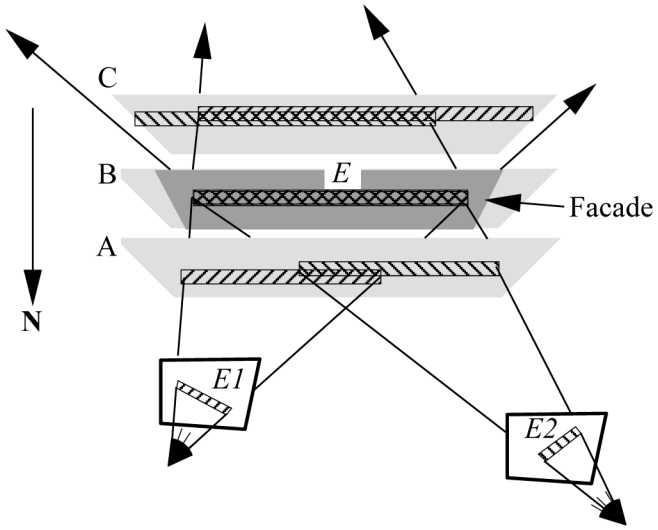

Figure 2.8: Image acquisition setup from controlled node locations (Coorg and Teller, 1999).

Coors *et al*. (2000) and Haala and Bohm (2003), both tried to simplify the problem of the dynamically moving camera with the use of direct referencing sensors to provide a coarse estimate of the extrinsic parameters. These sensors, which comprised a differential Global Positioning System (GPS) receiver, tilt sensors and a digital compass, were in addition to the digital camera and portable computer already carried by the user. The 3-D model of the target building was rendered based on the coarse estimates of the camera's parameters and a silhouette was generated as seen in Figure 2.9. This silhouette was overlaid on the terrestrial image and compared with the building edges previously extracted using the Generalized Hough Transform (Ballard and Brown, 1982). The location and orientation parameters of the camera were varied until a good alignment between the building outline on the terrestrial and the rendered silhouette was found, signifying a good estimate of the extrinsic parameters.

Figure 2.9: Silhouette of the building as projected to the image based on the extrinsic parameters obtained from GPS and digital compass (Haala and Bohm, 2003).

Debevec *et al*. (1996) demonstrated that the extrinsic parameters of multiple images can be recovered by minimizing a non-linear objective function which tabulates the error obtained between the observed line edge and the predicted line edge on the image. By using an initial estimate which assumes that the normal vector of the observed edge passes through the camera centre, they were able to accurately recover the location and orientation of the camera in an iterative manner.

After the camera parameters have been successfully recovered from the methods as described previously, the texture coordinates of each façade on the image can be calculated and façade texture automatically extracted. The subsequent assignment of textures to the building facades can be divided to two broad methods, namely using geometry dependent texture assignment and view dependent texture

assignment (Debevec *et al*., 1996; Coorg and Teller, 1999; Ortin and Remondino, 2005).

Geometry dependent texture assignment is a straight forward operation of assigning as the textures are registered according to building model space orientation hence automatic texturing is possible. On the contrary, view dependent texture assignment (Debevec *et al*., 1996; Coorg and Teller, 1999) required ray tracing calculations in order to produce seamless and occlusion-free façade textures from multiple images as demonstrated in Figure 2.10. This process was shown to be computationally intensive and needed to be processed off-line in order to merge the ray-traced projections to render a complete façade texture.

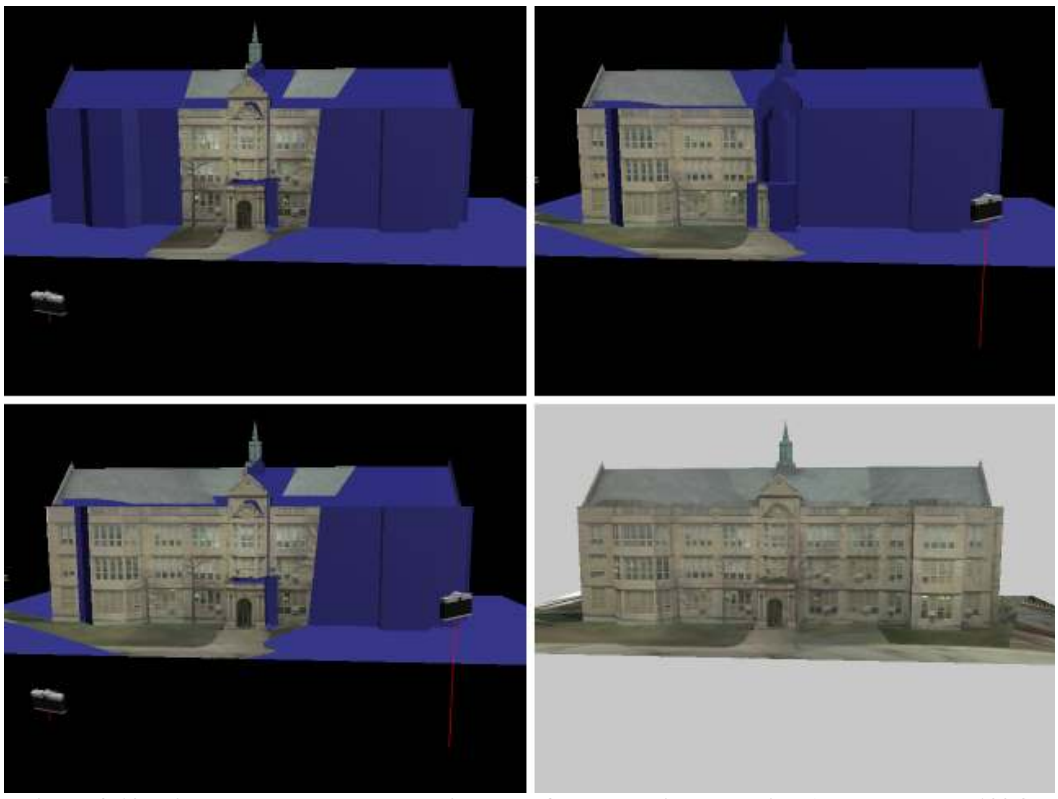

Figure 2.10: View dependent texture assignment for composite rendering (Debevec et al., 1996).

Chapter 3

Problem Statement

In the previous chapter, we reviewed the two broad methods used in the texture mapping of building models, namely Forward Texture Mapping (FTM) and Reverse Projection Texture Mapping (RPTM). In this chapter, we will examine the main motivation of this thesis, followed by the limitations of the current methods employed. The chapter will conclude by justifying the improvements of the proposed method over the current methods.

The main motivation of this thesis is to perform texture mapping using terrestrial digital photographs on standard 3-D building models in a quick and efficient manner. This is to be achieved without the need of direct referencing sensors for the estimation of extrinsic parameters belonging to the image acquisition station. Our definitions of the terms used in the above-mentioned are as follows:

1. "Standard 3-D building models" are defined as spatial data sets obtained from remote sensing stereo measurements.
2. "A quick and efficient manner" is defined as completing the entire texture mapping process on a single building model in 15 minutes or less. This includes the time taken for the extrinsic parameters recovery of the image acquisition station.
3. "Direct referencing sensors" are defined as GPS receivers, digital compasses and camera tilt/slant sensors.

3.1 Limitations of Current Methods

The limitations of the current methods used for texture mapping will be examined in this section, beginning with the Forward Texture Mapping (FTM) method followed by the Reverse Projection Texture Mapping (RPTM) method.

3.1.1 Forward Texture Mapping (FTM)

As described in Chapter 2, there are two broad methods for FTM, namely scene mosaics and polygon-based techniques.

Scene mosaicking is a technique which merges multiple images to form a scene composition. It has several limitations which severely reduces its efficiency for an otherwise elegant technique. Firstly, it requires the imaging equipment to be set up in a complicated manner during the acquisition stage. As described by Coorg and Teller (2000), images were acquired using a digital camera from a series of pre-determined tilt angles ranging from 50 to 70 degrees measured from the vertical axis at fixed node positions. At these node positions, the camera was mounted with a fixed optical centre on an indexed pan-tilt head attached to a tripod. As a result of the fixed positional nodes, they faced problems with sun flare as well as over-saturation of the imaging sensor during the image acquisition phase. The setup was dependent on weather conditions and would hamper the selection of an ideal location to acquire a good image of the subject.

Secondly, the process of merging of the multiple images requires the use of registration points between corresponding image pairs (Coorg and Teller, 2004; Zhang and Kang, 2004; Nicolas, 2001). These registration points are usually obtained via

correlation maximization functions performed in an iterative manner hence computationally intensive.

Thirdly, the scene mosaics are composed from images taken from the perspective view of each node and evidently retained its respective perspective characteristics. In order for the façade textures to be extracted and used effectively, explicit rectification of the scene mosaic needs to be done. Finally due to the size of resulting scene mosaics, they are not optimized for 3-D texture rendering as it is difficult to constrain the image sizes by powers of 2 (Wright and Lichak, 2005).

The polygon-based technique described by Tsai and Lin (2007) has a limitation which reduces the overall efficiency of the otherwise simple implementation. This method was labour intensive and time consuming as it required human assistance for a majority of its processing. Firstly, the identification of the façade AOI polygon on the image and the creation of respective enclosed polygons were performed manually by the user. Secondly, registration tie points had to be chosen between image pairs for the merging process of partial textures belonging to the same façade. Finally, the final mapping of the texture map to the building model was done individually for each polygon as the texture mosaic was registered to the object coordinate system of the building model. The entire process of performing FTM on a single building took an hour for a skilled operator. This did not include the time taken for the image acquisition. Furthermore, image registration needs to be done in a rigorous manner to avoid aliasing effects from being introduced to the final mosaic texture.

3.1.2 Reverse Projection Texture Mapping (RPTM)

RPTM can be divided into two processes, the initial camera location recovery phase followed by the subsequent texture assignment phase. There are three methods to

recover the camera's location for RPTM and two methods for texture assignment as described in Chapter 2. The camera's position can be recovered either by annotating the images with the absolute position and orientation from prior measurements or referencing using direct or indirect means. Texture assignment can be done using view-dependent or geometry-dependent mapping. There are limitations to the current methods of performing camera location recovery and texture assignment which reduce the overall efficiency of the work flow process.

Firstly, by using prior annotated images as demonstrated by Coorg and Teller (1999), precise measurements of the orientation and location need to be taken during the image acquisition stage. Furthermore, great care had to be taken in both the experimental setup phase and the annotation phase to ensure that error was reduced.

Secondly, for the usage of multiple direct referencing sensors for location recovery, it was shown by Haala and Bohm (2003) that such a technique could only provide a coarse estimate of the camera's location parameters. The equipment required was cumbersome as shown in Figure 3.1 and there were practical limitations to the accuracy of the direct referencing sensors. In the case of the GPS, its accuracy was limited especially in built-up areas as shadowing from high buildings could result in poor satellite coverage. Furthermore, signal reflections from nearby surrounding buildings could give rise to multi-path return effects to the GPS, which would further reduce the accuracy of the positional measurement. In the case of the digital compass which was supposed to provide the azimuth heading, Hoff and Azuma (2000) showed that such equipment was extremely vulnerable to distortions in built-up areas as well. They showed that the Earth's magnetic field could be influenced by cars and electrical installations, features which were commonly found in a city environment, reducing the overall accuracy of digital compasses to approximately 6 degrees. In order to refine the

accuracy of the recovered positional data, Haala and Bohm demonstrated a silhouette matching technique. However, this technique required a large search space in order to provide the best match as there were 6 degrees of freedom involved in the location and orientation parameters. It was computationally intensive and had to be performed off-line due to the number of iterations involved.

Figure 3.1: Prototype of direct referencing sensors rig (Haala and Bohm, 2003).

Moving on to the texture assignment process, view-dependent texture mapping was shown by Debevec *et al*. (1996) to create seamless facade textures which were representative of the building model acquired from multiple view angles. The drawback was the computational resources required for view-dependent texture mapping as ray tracing calculations were done for every individual pixel on the facade texture to obtain the correct pixel value. Furthermore, this technique was effective only if the building model conformed very closely to the actual building structure. This is

not the case for almost all available 3-D city model data sets since currently the most efficient and feasible way of performing area 3-D collection of these models is via airborne sensors. The drawback of using airborne sensors is that the amount of structural details belonging to the building façade which can be extracted is very limited. Debevec *et al*. were able to avoid mismatch ambiguities between low detail building model data sets and high resolution ground images by generating their own 3-D building model structures using human assisted stereo reconstruction. This was done using the same set of terrestrial images as a prior step to view-dependent texture assignment. It must be noted that human assisted stereo reconstruction is both time consuming and laborious, depending a fair bit on the operator's skill and experience.

3.2 Improvements over Current Methods

In this section, we will examine the factors whereby the current methods of performing texture mapping on standard building models do not fulfill the requirements of the problem statement.

Firstly, FTM despite being a simpler method compared to RPTM is significantly slower and more labour intensive. This is because FTM requires a large amount of human assistance for control point selection and polygon AOI creation. Secondly, current RPTM methods are more efficient and streamlined compared to FTM. The main drawback is that they require either direct referencing sensors or highly accurate and detailed 3-D building models. We have seen in the previous section that current generation of direct referencing sensors are unable to provide the desired accuracy required for texture mapping purposes. Furthermore, highly detailed facades are not a standard feature on commonly available 3-D data sets as the feature extraction process is currently impractical. This drastically reduces the efficiency of automatic feature matching using correlation between the image and model facade which is often used by RPTM methods.

We propose a method of performing an improved version of RPTM without the need for direct referencing sensors and highly detailed building model data sets. The improved RPTM method will determine the camera's location using a hybrid process to assist in the reverse projection of the facade polygon outlines onto the digital photograph. The alignment of the AOI polygon outlines with the façade will re-affirm the accuracy of the camera’s extrinsic parameters for texture extraction. Some human assistance is required to determine the level of accuracy of the building data before the hybrid process of image location recovery can take place. If the building model features are similar to the terrestrial photograph, an improved space resection

technique will be used to recover the camera's location; otherwise a pose manipulation approach will be used instead.

Chapter 4

Methodology

This chapter describes in detail the methods used in the implementation of the Improved Reverse Projection Texture Mapping (IRPTM) without the need of direct referencing sensors and highly detailed building model data sets. The overview of the building texturing process can be seen in Figure 4.1. A program named Graffiti was implemented using IRPTM for the texturing and visual interpretation of building models in a mixed reality environment. Graffiti has 4 main stages which perform different functions. The initial data input stage takes in unregistered terrestrial photographs, satellite map images and standard 3-D building models. Next, the correct façade textures are extracted in the façade extraction stage. IRPTM is most involved in this stage and it forms the main core of the Graffiti program. This is followed by the data fusion stage which merges the extracted textures with the respective building models. Finally, the user is able to view the final integrated building product in the data visualization stage with the option to save and export his/her work.

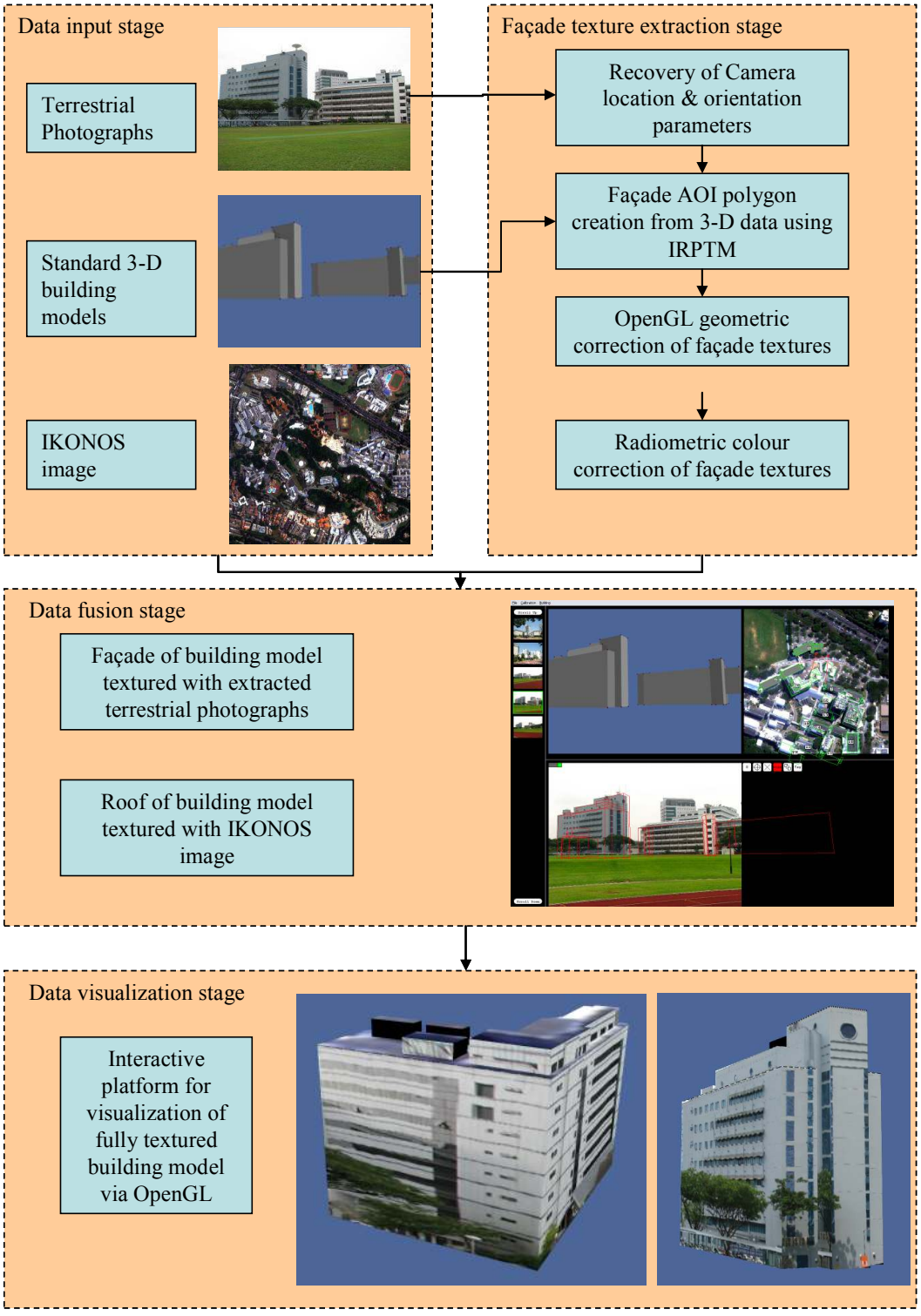

Figure 4.1: Work flow of implemented Graffiti program.

Texture mapping of building façades involve the transformation of pixel information from image space to building model space. In our IRPTM implementation as shown in Figure 4.2, four 3-D points (*X, Y, Z*) belonging to a single façade coupled with the intrinsic and extrinsic parameters of the acquisition camera are transformed to four 2-D (*x, y*) points on the terrestrial image using the following collinear equations:

$$x_n = x_0 - f\left[\frac{m_{11}(X_n - X_{camera}) + m_{12}(Y_n - Y_{camera}) + m_{13}(Z_n - Z_{camera})}{m_{31}(X_n - X_{camera}) + m_{32}(Y_n - Y_{camera}) + m_{33}(Z_n - Z_{camera})}\right] \quad (4.1)$$

$$y_n = y_0 - f\left[\frac{m_{21}(X_n - X_{camera}) + m_{22}(Y_n - Y_{camera}) + m_{23}(Z_n - Z_{camera})}{m_{31}(X_n - X_{camera}) + m_{32}(Y_n - Y_{camera}) + m_{33}(Z_n - Z_{camera})}\right] \quad (4.2)$$

where x_n, y_n are the image coordinates of the transformed facade
X_n, Y_n, Z_n are the object coordinates of the facade
$X_{camera}, Y_{camera}, Z_{camera}$ are the object coordinates of the camera
$m_{11} - m_{33}$ are the components of the orientation matrix
f is the focal length
x_0, y_0 are the image coordinates of the image centre
$n = 1, 2, 3, 4$

The resulting four 2-D points will form an Area of Interest (AOI) polygon on the terrestrial image from which the façade can extract textural pixel information. This is done for all the façades belonging to the 3-D building. As seen from Figure 4.2, the accuracy of the camera's location and orientation is essential for the façade texture to be recovered. Conversely, it is an interesting point to note that when the camera's extrinsic parameters are correct, the AOI polygons will be perfectly aligned with their real-world counterparts in the terrestrial photograph. This will be used in a novel manner to verify our answers obtained in the evaluation section later in Chapter 5.

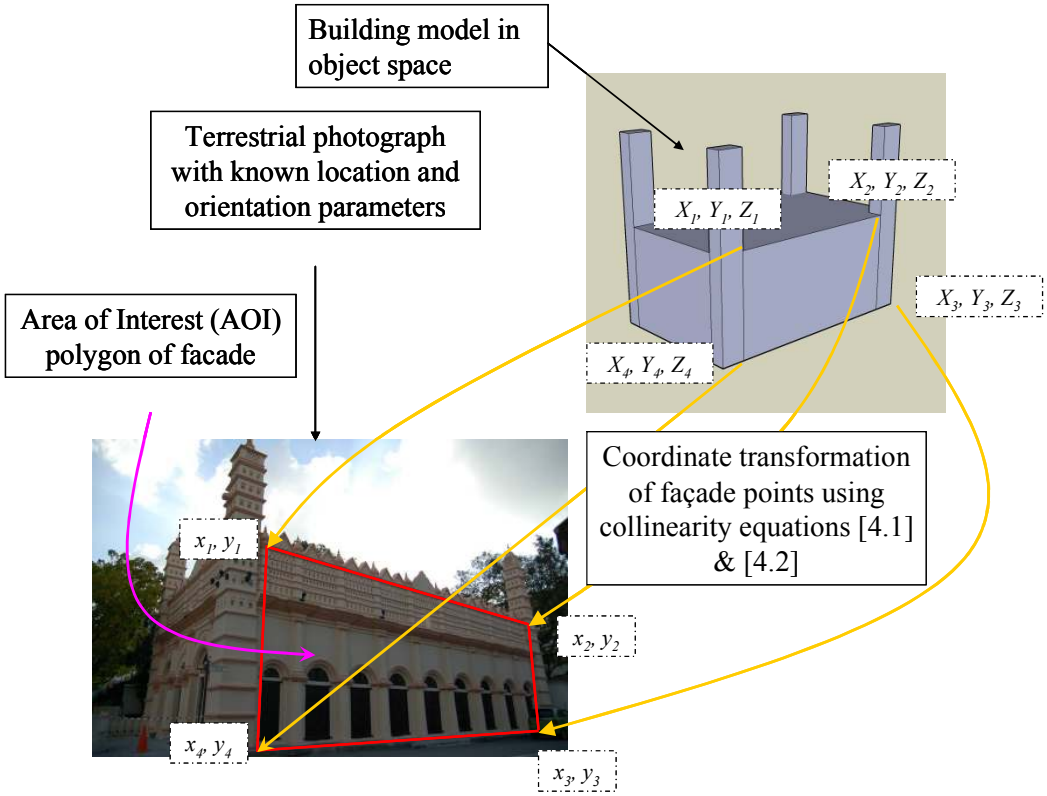

Figure 4.2: Reverse Projection Texture Mapping.

The methods which were used in the implementation of IRPTM in the façade texture extraction stage are as follows:

i. Determination of Camera Extrinsic Parameters (I) – A hybrid space resection method.

ii. Determination of Camera Extrinsic Parameters (II) – An interactive pose estimation method.

iii. Geometric Correction of Skewed Façade Textures.

iv. Colour Balancing of Extracted Façade Textures.

v. Mosaic Composition of Multiple Images.

4.1 Determination of Camera Extrinsic Parameters (I) – A Hybrid Space Resection Method

As seen in Chapter 3, the recovery of the extrinsic parameters of the camera is imperative for the accurate projection of building façade AOI polygons onto the terrestrial photograph. Space resection is the process by which the extrinsic parameters of the camera are determined based on the measurements of the images of ground control points appearing on the photograph.

4.1.1 Iterative Space Resection

Without relying on external direct referencing sensors, the extrinsic parameters can be recovered accurately using space resection by collinearity in an iterative manner (Wolf and Dewitt, 2000). Rewriting the collinearity equations in Eqs. (4.1) and (4.2), we have:

$$F = x_0 - f\frac{R}{Q} = x_n \tag{4.3}$$

$$G = y_0 - f\frac{S}{Q} = y_n \tag{4.4}$$

$$Q = m_{31}\left(X_n - X_{camera}\right) + m_{32}\left(Y_n - Y_{camera}\right) + m_{33}\left(Z_n - Z_{camera}\right) \tag{4.5}$$

$$R = m_{11}\left(X_n - X_{camera}\right) + m_{12}\left(Y_n - Y_{camera}\right) + m_{13}\left(Z_n - Z_{camera}\right) \tag{4.6}$$

$$S = m_{21}\left(X_n - X_{camera}\right) + m_{22}\left(Y_n - Y_{camera}\right) + m_{23}\left(Z_n - Z_{camera}\right) \tag{4.7}$$

According to Taylor's theorem, Eqs. (4.3) and (4.4) can be expressed in linearized form by taking the partial derivatives with respect to the unknown variables, φ, ω, κ, X_{camera}, Y_{camera} and Z_{camera} as shown in Eqs. (4.8) and (4.9):

$$F_0 + \left(\frac{\partial F}{\partial \omega}\right)_0 \partial\omega + \left(\frac{\partial F}{\partial \varphi}\right)_0 \partial\varphi + \left(\frac{\partial F}{\partial \kappa}\right)_0 \partial\kappa$$
$$+ \left(\frac{\partial F}{\partial X_{camera}}\right)_0 \partial X_{camera} + \left(\frac{\partial F}{\partial Y_{camera}}\right)_0 \partial Y_{camera} + \left(\frac{\partial F}{\partial Z_{camera}}\right)_0 \partial Z_{camera} \qquad (4.8)$$
$$+ \left(\frac{\partial F}{\partial X_n}\right)_0 \partial X_n + \left(\frac{\partial F}{\partial Y_n}\right)_0 \partial Y_n + \left(\frac{\partial F}{\partial Z_n}\right)_0 \partial Z_n = x_n$$

$$G_0 + \left(\frac{\partial G}{\partial \omega}\right)_0 \partial\omega + \left(\frac{\partial G}{\partial \varphi}\right)_0 \partial\varphi + \left(\frac{\partial G}{\partial \kappa}\right)_0 \partial\kappa$$
$$+ \left(\frac{\partial G}{\partial X_{camera}}\right)_0 \partial X_{camera} + \left(\frac{\partial G}{\partial Y_{camera}}\right)_0 \partial Y_{camera} + \left(\frac{\partial G}{\partial Z_{camera}}\right)_0 \partial Z_{camera} \qquad (4.9)$$
$$+ \left(\frac{\partial G}{\partial X_n}\right)_0 \partial X_n + \left(\frac{\partial G}{\partial Y_n}\right)_0 \partial Y_n + \left(\frac{\partial G}{\partial Z_n}\right)_0 \partial Z_n = y_n$$

Eqs. (4.8) and (4.9) can be further simplified into Eqs. (4.10) and (4.11) for usage in a least squares adjustment system to reduce the errors between the observed value and true value. Variables $d\varphi$, $d\omega$, $d\kappa$, dX_{camera}, dY_{camera} and dZ_{camera} are the unknown corrections applied to the initial approximations of φ, ω, κ, X_{camera}, Y_{camera} and Z_{camera}. The ground control points X_n, Y_n, Z_n (n = 1- 4) are assumed to be known and static hence the values of ∂X_n, ∂Y_n, ∂Z_n are zero.

$$b_{11}d\omega + b_{12}d\varphi + b_{13}d\kappa - b_{14}dX_{camera} - b_{15}dY_{camera} - b_{16}dZ_{camera} = x_n - x_0 + f\frac{R}{Q} \qquad (4.10)$$

$$b_{21}d\omega + b_{22}d\varphi + b_{23}d\kappa - b_{24}dX_{camera} - b_{25}dY_{camera} - b_{26}dZ_{camera} = y_n - y_0 + f\frac{S}{Q} \qquad (4.11)$$

where

$$b_{11} = \frac{f}{Q^2}\left[R\left(-m_{33}\Delta Y + m_{32}\Delta Z\right) - Q\left(-m_{13}\Delta Y + m_{12}\Delta Z\right)\right]$$

$$b_{12} = \frac{f}{Q^2}\left[R\left(\cos\varphi\,\Delta X + \sin\omega\sin\varphi\,\Delta Y - \cos\omega\sin\varphi\,\Delta Z\right)\right.$$
$$\left. - Q\left(-\sin\varphi\cos\kappa\,\Delta X + \sin\omega\cos\varphi\cos\kappa\,\Delta Y - \cos\omega\cos\varphi\cos\kappa\,\Delta Z\right)\right]$$

$$b_{13} = \frac{-f}{Q}\left(m_{21}\Delta X + m_{22}\Delta Y + m_{23}\Delta Z\right)$$

$$b_{14} = \frac{f}{Q^2}\left(Rm_{31} - Qm_{11}\right)$$

$$b_{15} = \frac{f}{Q^2}\left(Rm_{32} - Qm_{12}\right)$$

$$b_{16} = \frac{f}{Q^2}\left(Rm_{33} - Qm_{13}\right)$$

$$b_{21} = \frac{f}{Q^2}\left[S\left(-m_{33}\Delta Y + m_{32}\Delta Z\right) - Q\left(-m_{23}\Delta Y + m_{22}\Delta Z\right)\right]$$

$$b_{22} = \frac{f}{Q^2}\left[S\left(\cos\varphi\,\Delta X + \sin\omega\sin\varphi\,\Delta Y - \cos\omega\sin\varphi\,\Delta Z\right)\right.$$
$$\left. - Q\left(\sin\varphi\sin\kappa\,\Delta X - \sin\omega\cos\varphi\sin\kappa\,\Delta Y + \cos\omega\cos\varphi\sin\kappa\,\Delta Z\right)\right]$$

$$b_{23} = \frac{f}{Q}\left(m_{11}\Delta X + m_{12}\Delta Y + m_{13}\Delta Z\right)$$

$$b_{24} = \frac{f}{Q^2}\left(Sm_{31} - Qm_{21}\right)$$

$$b_{25} = \frac{f}{Q^2}\left(Sm_{32} - Qm_{22}\right)$$

$$b_{26} = \frac{f}{Q^2}\left(Sm_{33} - Qm_{23}\right)$$

$$\Delta X = X_n - X_{camera}$$

$$\Delta Y = Y_n - Y_{camera}$$

$$\Delta Z = Z_n - Z_{camera}$$

From Eqs. (4.10) and (4.11), a least squares solution can be computed using at least 4 sets of ground control points matched with their corresponding points on the image. The choice of the initial estimate for φ, ω, κ, X_{camera}, Y_{camera} and Z_{camera} is vital as the solution might converge erroneously due to the local minima encountered.

4.1.2 Closed-form Space Resection

The closed-form space resection method (Zeng and Wang, 1992) does not require any initial estimates or iterations in the computation of the extrinsic parameters. As shown in Figures 4.3 and 4.4, this method utilizes only 3 sets of ground control points and their corresponding image control points in order to set up Eqs. (4.12) and (4.13):

$$\begin{aligned} x_1 &= -f \times \frac{X_1}{Z_1}, \quad y_1 = -f \times \frac{Y_1}{Z_1} \\ x_2 &= -f \times \frac{X_2}{Z_2}, \quad y_2 = -f \times \frac{Y_2}{Z_2} \\ x_3 &= -f \times \frac{X_3}{Z_3}, \quad y_3 = -f \times \frac{Y_3}{Z_3} \end{aligned} \tag{4.12}$$

$$\begin{aligned} (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 &= L_3^2 \\ (X_2 - X_3)^2 + (Y_2 - Y_3)^2 + (Z_2 - Z_3)^2 &= L_1^2 \\ (X_3 - X_1)^2 + (Y_3 - Y_1)^2 + (Z_3 - Z_1)^2 &= L_2^2 \end{aligned} \tag{4.13}$$

$$\begin{aligned} (X_1 - X_{camera})^2 + (Y_1 - Y_{camera})^2 + (Z_1 - Z_{camera})^2 &= R_1^2 \\ (X_2 - X_{camera})^2 + (Y_2 - Y_{camera})^2 + (Z_2 - Z_{camera})^2 &= R_2^2 \\ (X_3 - X_{camera})^2 + (Y_3 - Y_{camera})^2 + (Z_3 - Z_{camera})^2 &= R_3^2 \end{aligned} \tag{4.14}$$

Possible solutions for the extrinsic parameters φ, ω, κ, X_{camera}, Y_{camera} and Z_{camera} can be recovered algebraically by finding the intersection point of three spheres with radius R_1, R_2 and R_3 in Eq. (4.14). An additional 4th ground control point is required to verify the solution to Eq. (4.14) as it can return up to four possible orientation solutions. The main drawbacks are that this method is error intolerant and it does not use the 4th control point for actual calculation for the extrinsic parameters of the exposure station.

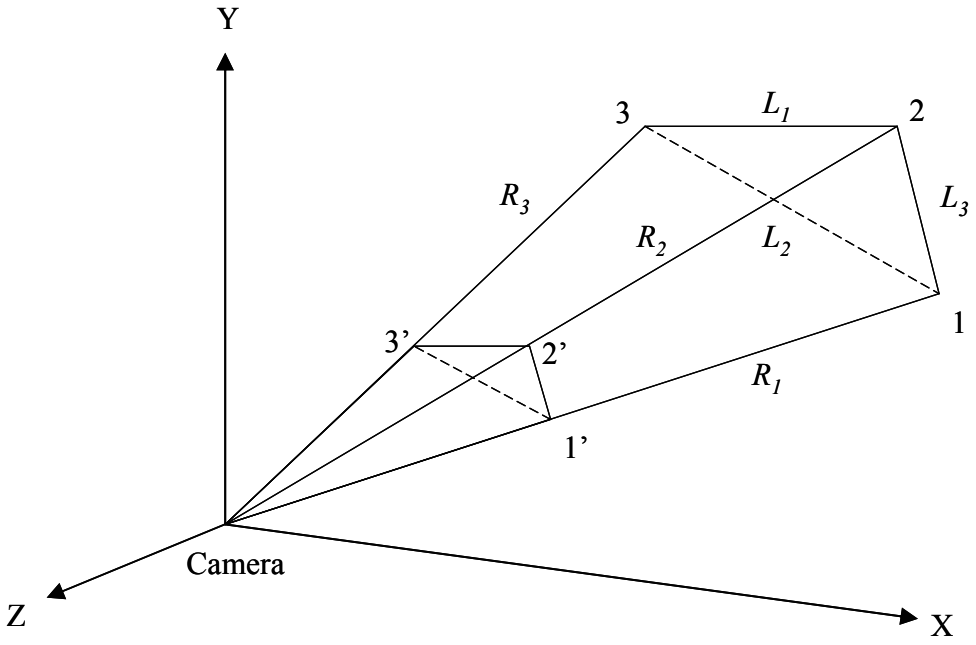

Figure 4.3: Image pyramid (Camera is at origin; 1, 2, 3 are object control points).

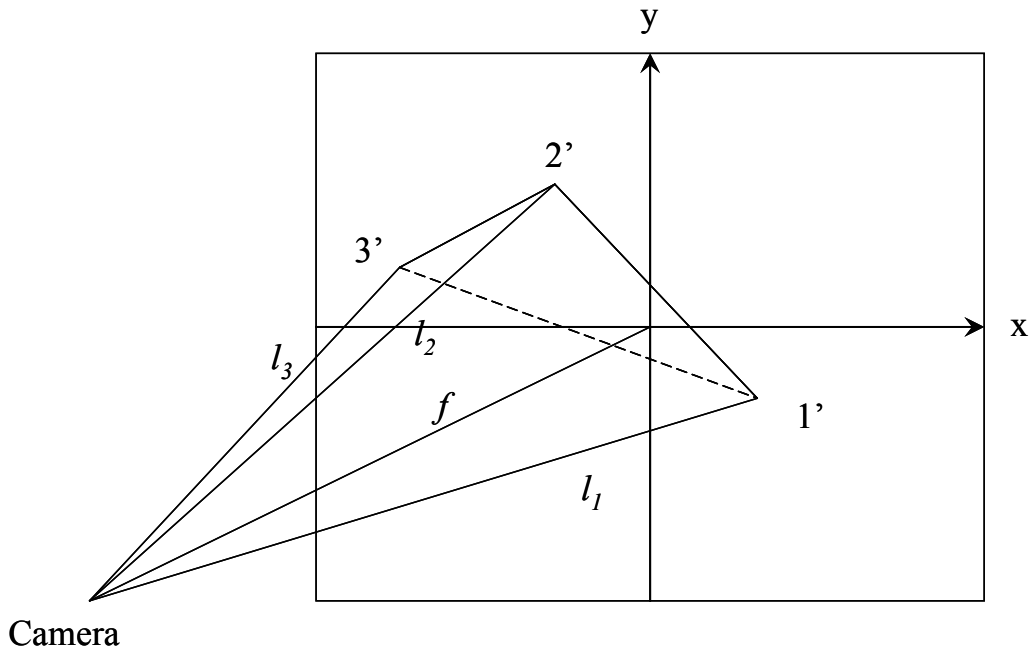

Figure 4.4: The geometry of an image plane and camera.

4.1.3 Hybrid Space Resection

The hybrid space resection method which we adopt utilizes both the closed-form space resection as well as the iterative space resection in the recovery of the extrinsic parameters. Since both the above mentioned methods require a minimum of 4 ground control points, we are able to avoid the pitfalls of both methods by fully utilizing the 4 sets of control points for our calculations. The solution of a closed-form space resection method is used as the initial estimation and further integrated by using all 4 control points in the iterative space resection method. The flowchart of this process is illustrated more clearly in Figure 4.5. As such, we are able to reduce the possibility of the iterative solution converging to local minima and maximize the utility of all the 4 ground control points.

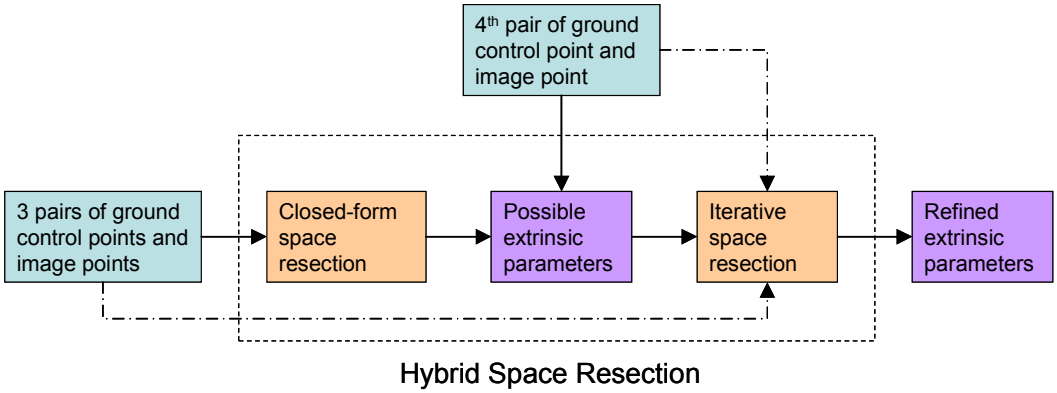

Figure 4.5: Flowchart of Hybrid Space Resection process.

4.2 Determination of Camera Extrinsic Parameters (II) – An interactive pose estimation method

OpenGL is a standard specification for writing applications that produce 2-D and 3-D computer graphics, used widely in the field of visualization. The OpenGL camera is the vantage point in the virtual world whereby the user is given a view of his surroundings. This rendered view is very much dependent on the position and orientation of the OpenGL camera with respect to the surrounding objects and this is known as the pose. It is similar for the real world camera whereby it captures a composition of multiple real objects in an image based on its pose.

We can make use of OpenGL for pose estimation in our application as the rendered scene of the 3-D building models is similar to that of a photograph taken by a perspective camera. The OpenGL camera and the real perspective camera share similarities with respect to their intrinsic and extrinsic parameters, hence pose estimation can be done in the following steps:

1. Setup of OpenGL camera with real camera intrinsic parameters.
2. Varying of position and orientation of OpenGL camera.
3. Reverse projection of façade AOI polygon outlines onto the terrestrial photograph using the extrinsic parameters obtained from step 2.
4. Check similarity match between projected façade AOI polygon and actual building on the terrestrial photograph.

The rendered scene of the OpenGL camera is highly dependent on the field of view angles in the horizontal and vertical directions as denoted by Eqs. (4.15) and (4.16). As shown in Figure 4.6, the OpenGL camera's parameters can be initialized according to the intrinsic parameters of a real camera. Using Eq. (4.17), we can recover the vertical

field of view angle in the real camera if it is modeled after an ideal pinhole camera as shown in Figure 4.7.

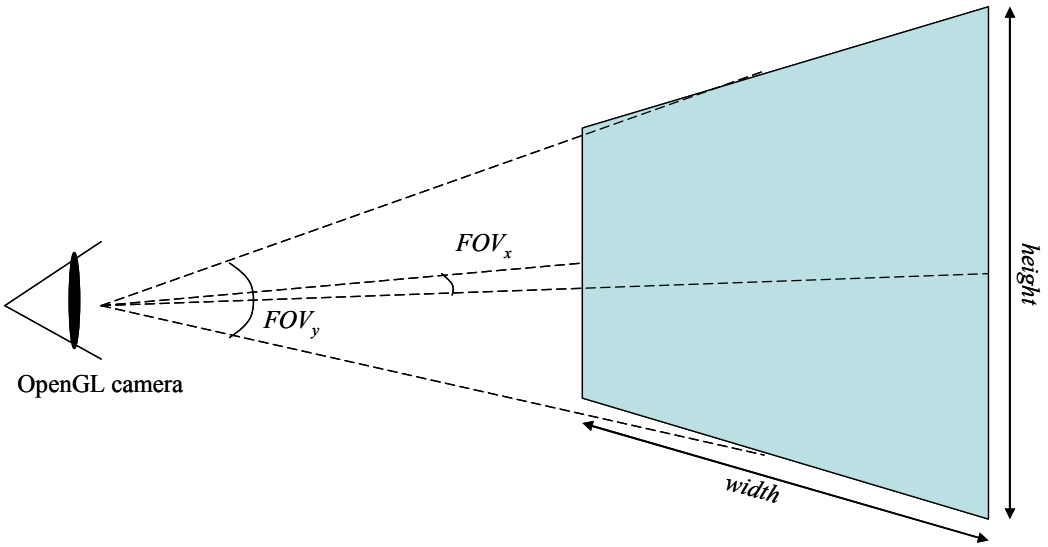

Figure 4.6: Relationship between field of view angle in the x and y direction for an OpenGL camera.

$$aspect = \frac{width}{height} \tag{4.15}$$

$$FOV_x = FOV_y \times aspect \tag{4.16}$$

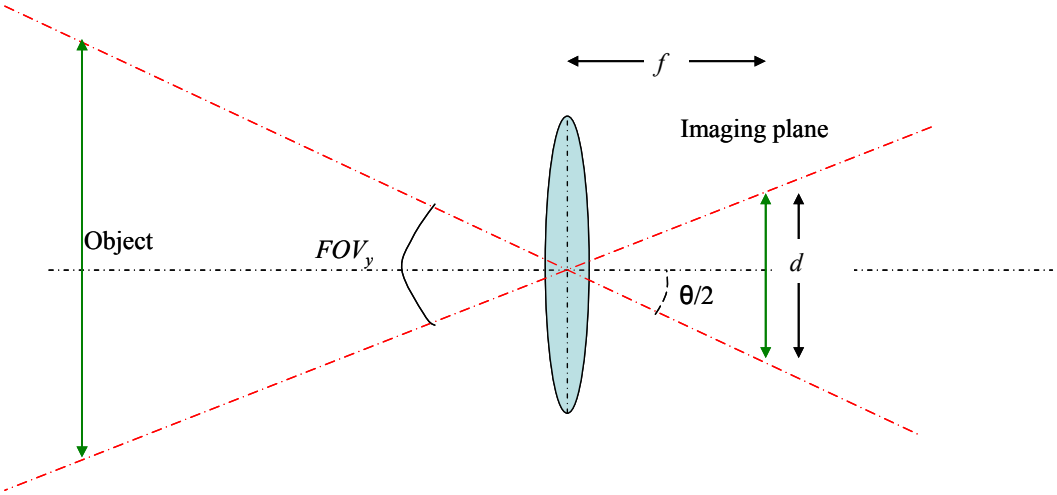

Figure 4.7: Relationship between field of view angle in the y direction, focal length and imaging plane in a pin hole camera model.

$$FOV_y = 2\arctan\left(\frac{d}{2f}\right) \tag{4.17}$$

By varying the position and orientation of the OpenGL camera, the user can adjust the rendered scene to match the pose of the terrestrial photograph. The closeness of match can be reaffirmed by performing IRPTM using Eqs. (4.1) and (4.2) on the building model to project the façade AOI polygons onto the terrestrial photograph using the extrinsic parameters obtained from the OpenGL camera. The 3-D building model used for the matching process does not need to be highly detailed as the user can decide from the rendered pose whether the OpenGL camera's pose fit those of the actual camera. When a close match has been established between the OpenGL rendered scene and the terrestrial photograph, the latter is able to use the pose information of the OpenGL camera as its own extrinsic parameters. We are able to further refine the accuracy of the recovered pose by using it as the initial estimate and passing through the iterative space resection method described in Section 4.1.1 with 4 additional ground control points.

4.3 Geometric Correction of Skewed Façade Textures

In the case of photographing buildings, an oblique view would be defined as when the normal of the camera's imaging plane is not parallel to the normal belonging to the target building façade. Conversely, a normal view can be defined as when the normal of the camera's imaging plane is parallel to that of the normal belonging to the target building façade. Oblique view terrestrial photographs of the buildings are more effective in large scale texture mapping as compared to frontal normal view photographs.

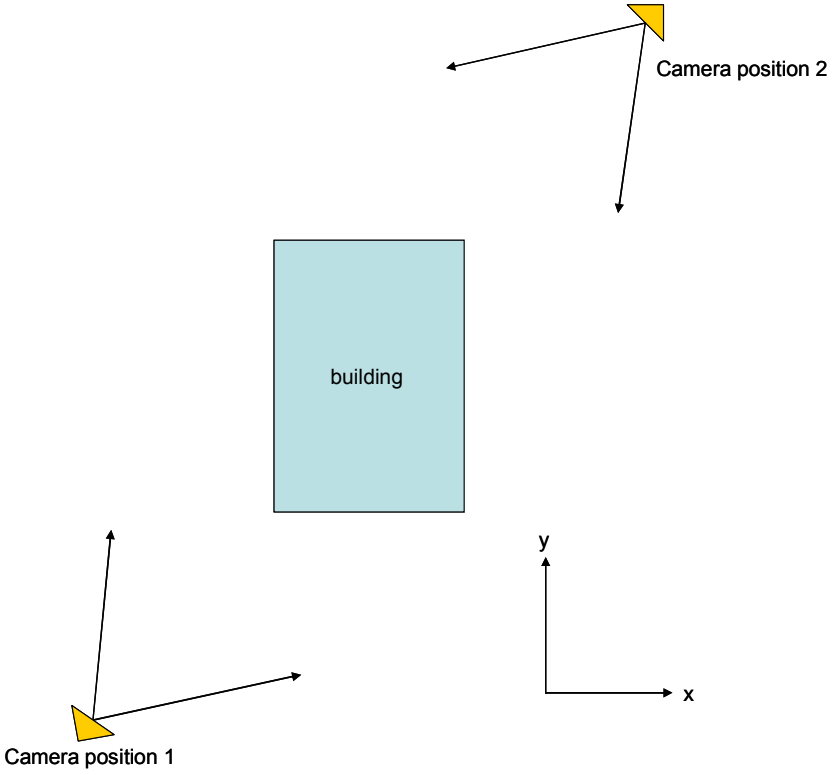

Figure 4.8: Photographing a building in an oblique view manner.

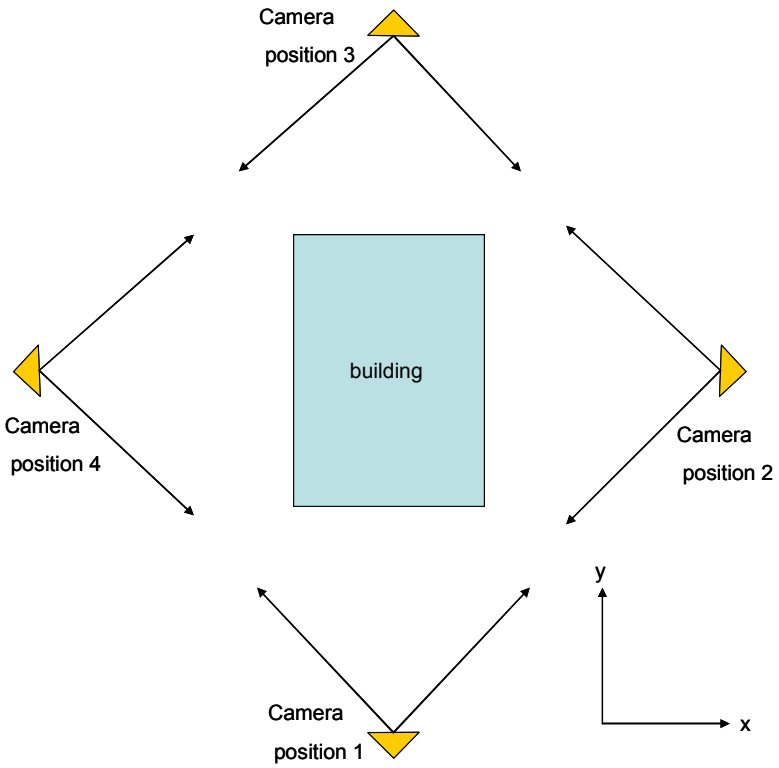

Figure 4.9: Photographing a building in a normal view manner.

As seen in Figures 4.8 and 4.9, oblique view imaging is able to capture more views of the building façade with lesser images as compared to normal view imaging. The oblique view method required 2 image acquisition positions for a 4 sided building as compared with the normal view method which required 4 image acquisition positions. As a result of using oblique view photographs of the building façade, a skew factor will be introduced into the extracted texture and it is highly dependent on the angle of which the photograph was taken with respect to the building. This is due to the perspective characteristics of the camera used. We are able to correct this geometric distortion using a method devised in OpenGL as shown in Figure 4.10. By projecting the extracted façade texture onto a regular OpenGL quadrilateral using the texture coordinates obtained using Eq. (4.18) for the corner vertices, the skewed texture can be geometrically corrected in real-time.

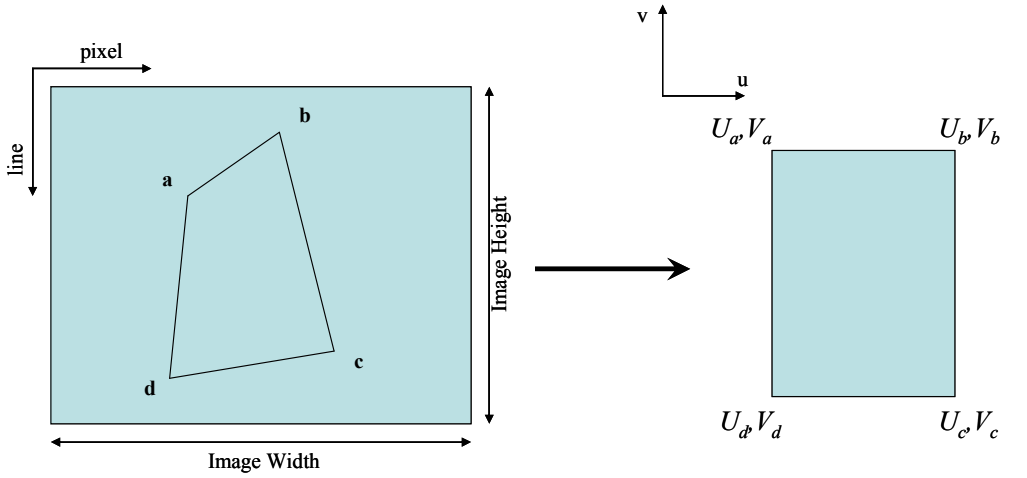

Figure 4.10: Geometric correction of a skewed façade texture using OpenGL.

$$U_n = \frac{pixel_n}{image\ width}$$
$$V_n = 1 - \frac{line_n}{image\ height} \qquad (4.18)$$
$$n = a, b, c, d$$

4.4 Colour Balancing of Extracted Façade Textures

After texture mapping is carried out, the resulting building model would consist of a mosaic of different façade textures acquired from different angles and physical conditions. This results in differences in colour and contrast on different sides of the building model. We can produce a seamless transition effect between different façade textures by performing histogram matching using statistical methods (Gonzales and Woods, 2002) to reduce the differences. As shown in Eq. (4.19), the three colour values of each pixel are normalized according to its image's respective mean and standard deviation, followed by a scaling and offset to match the statistics that of the reference image.

$$
\begin{aligned}
p_{x,y}^{Balanced,red} &= \left(p_{x,y}^{original,red} - \mu^{original,red}\right) \times \frac{\sigma^{reference,red}}{\sigma^{original,red}} + \mu^{refernce,red} \\
p_{x,y}^{Balanced,green} &= \left(p_{x,y}^{original,green} - \mu^{original,green}\right) \times \frac{\sigma^{reference,green}}{\sigma^{original,green}} + \mu^{refernce,green} \\
p_{x,y}^{Balanced,blue} &= \left(p_{x,y}^{original,blue} - \mu^{original,blue}\right) \times \frac{\sigma^{reference,blue}}{\sigma^{original,blue}} + \mu^{refernce,blue}
\end{aligned}
\tag{4.19}
$$

4.5 Mosaic Composition of Multiple Images

It is possible that certain building model facades cannot be fully textured by a single terrestrial image as the façade width might be greater than what the camera can physically capture. We can overcome this limitation by creating a mosaic composition when more than one image is used to photograph the entire facade. The relationship of any point in a projective plane can be mapped using a unique projective transformation as shown in the following equation:

$$
\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4.20}
$$

By re-arranging Eq. (4.20), we obtain Eq. (4.21) using 4 unique non-collinear and corresponding control points on the terrestrial image as shown below. This is subsequently solved by minimizing the residuals and we can obtain the plane projective transformation matrix:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2'x_2 & -x_2'y_2 & -x_2' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y_2'x_2 & -y_2'y_2 & -y_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3'x_3 & -x_3'y_3 & -x_3' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y_3'x_3 & -y_3'y_3 & -y_3' \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4'x_4 & -x_4'y_4 & -x_4' \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y_4'x_4 & -y_4'y_4 & -y_4' \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = 0 \qquad (4.21)$$

By using the transformation matrix, we are able to extend the boundaries of the base image onto the next corresponding image, hence widening the total viewing area of the photograph.

Chapter 5

Evaluation of Improved Reverse Projection Texture Mapping

IRPTM will be used to generate façade textures for the standard 3-D building models using terrestrial photographs. Firstly, the accuracy of both the hybrid space resection as well as the pose estimation method will be evaluated for the recovery of the camera's extrinsic parameters. Secondly, the quality of the extracted façade texture will be examined. This is evaluated in terms of the geometric accuracy of the reproduced façade texture as well as the seamless continuity of the façade texture mosaic.

5.1 Evaluation of Camera Extrinsic Parameters Recovery Methods

As discussed earlier in Chapter 4, the extrinsic parameters are important variables which denote the location and orientation at the instance the terrestrial photograph was taken. These are vital for accurate texture extraction as they will affect the formation of the façade AOI polygon after IRPTM is carried out.

5.1.1 Accuracy of Recovered Extrinsic Parameters Using Hybrid Space Resection

The extrinsic parameters of the image were recovered using 3 different sets of test data which can be found in Appendix A (Kwoh, 1988; Wolf and Dewitt, 2000). Each data

set comprises of 4 pairs of pre-determined ground control points and corresponding image points. A computer program was implemented using both methods described in Sections 4.1.2 and 4.1.3 and used to recover the location and orientation of the camera during acquisition.

The recovered extrinsic parameters are shown in Tables 5.1 – 5.3 and the Euclidean distance between the actual camera parameters and the recovered parameters can be seen in Table 5.4. It was observed that the recovered extrinsic parameters for both closed-form resection and the hybrid resection were close to the actual parameters of the camera. However, the deviation of the camera location from the actual location using the hybrid resection method was smaller compared to the closed-formed space resection method, with an average difference of less than 0.7 meter.

Table 5.1: Recovered extrinsic parameters using Closed-form Space Resection and Hybrid Space Resection for data set 1.

Extrinsic Parameters	Actual Position	Closed-form Space Resection	Hybrid Space Resection
X / meters	1027.86	1027.82	1027.86
Y / meters	1044.11	1044.10	1044.15
Z / meters	648.20	648.21	648.18
φ / radian	-0.00717	-0.00716	-0.00722
ω / radian	0.02112	0.02106	0.0212
κ / radian	1.7942	1.79419	1.79421

Table 5.2: Recovered extrinsic parameters using Closed-form Space Resection and Hybrid Space Resection for data set 2.

Extrinsic Parameters	Actual Position	Closed-form Space Resection	Hybrid Space Resection
X / meters	9157.27	9157.14	9157.74
Y / meters	17864.07	17865.39	17864.48
Z / meters	1562.13	1561.60	1562.01
φ / radian	0.01429	0.01212	0.01267
ω / radian	0.02535	0.02285	0.02291
κ / radian	-1.79624	-1.83904	-1.83915

Table 5.3: Recovered extrinsic parameters using Closed-form Space Resection and Hybrid Space Resection for data set 3.

Extrinsic Parameters	Actual Position	Closed-form Space Resection	Hybrid Space Resection
X / meters	8935.93	8937.03	8935.41
Y / meters	17032.27	17032.60	17032.31
Z / meters	1581.52	1581.56	1581.72
φ / radian	-0.00160	-0.00178	-0.00157
ω / radian	0.03991	0.03654	0.03570
κ / radian	-1.79768	-1.82434	-1.82444

Table 5.4: Euclidean distance between the actual camera's location and the recovered camera's location.

	Euclidean distance (Closed-form Recovery) / meters	Euclidean distance (Hybrid Recovery) / meters
Data Set 1	0.04	0.04
Data Set 2	1.15	0.56
Data Set 3	1.43	0.63

As seen in Tables 5.5 and 5.6, we can introduce some Gaussian noise to the control points in varying degree to evaluate the robustness of the hybrid space resection method against measurement errors. The noise added to the image points has 0 mean and a standard deviation of ± 0.10 mm and the noise added to the ground control points has 0 mean and a standard deviation of ± 1.00 meters. It was observed that closed-form space resection was more sensitive to noise as compared to hybrid resection. Furthermore, closed-form resection recovered camera locations showed a greater deviation when noise was added to the ground control points as compared to the image points. Hybrid space resection recovered camera locations was observed to have overall smaller deviations when using both noisy ground control points and image points as compared to closed-form space resection.

Table 5.5: Euclidean distance between actual camera and recovered camera when noise was added to image points (data set 3).

±0.1mm variation in n image points	Euclidean distance (Hybrid recovery) / meters	Euclidean distance (Closed-form recovery) / meters
n = 1	3.59	5.79
n = 2	4.14	8.96
n = 3	2.48	7.67
n = 4	3.95	7.67

Table 5.6: Euclidean distance between actual camera and recovered camera when noise was added to ground control points (data set 3).

± 1.0m variation in n ground control point	Euclidean distance (Hybrid recovery) / meters	Euclidean distance (Closed-form recovery) / meters
n = 1	3.08	6.45
n = 2	8.24	10.97
n = 3	3.19	20.54
n = 4	4.45	20.54

5.1.2 Accuracy of Recovered Extrinsic Parameters Using Interactive Pose Estimation

Using the Graffiti program which was described previously in Chapter 4, we will attempt to recover the camera's extrinsic parameters using the pose estimation method on standard 3-D models and digital terrestrial photographs. As shown earlier in Section 5.1.1, hybrid space resection method was able to recover the extrinsic parameters of the camera accurately, hence it will be used as a yard stick to validate and gauge the performance of pose estimation.

The recovered camera extrinsic parameters using both the hybrid space resection method and the pose estimation method on terrestrial photographs taken by different cameras can be seen in Tables 5.7 – 5.12 and the resulting Euclidean distance between them can be observed in Table 5.13. It was observed that the pose estimated extrinsic parameters were very close to those recovered by hybrid space resection method. The overall average difference is about 1.7 meters for the 3 cameras used.

Table 5.7: Recovered camera extrinsic parameters for test image dsc_5403.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364308.48	364308.08
Y / meters	143447.99	143447.92
Z / meters	60.42	59.43
φ / radians	-1.359	-1.364
ω / radians	-0.034	-0.037
κ / radians	-3.127	-3.129

Table 5.8: Recovered camera extrinsic parameters for test image dsc_7967.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364439.04	364439.26
Y / meters	143443.02	143443.96
Z / meters	14.62	15.26
φ / radians	-1.772	-1.723
ω / radians	0.855	0.846
κ / radians	-2.995	-3.033

Table 5.9: Recovered camera extrinsic parameters for test image img_7089.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364455.31	364454.90
Y / meters	143438.47	143439.07
Z / meters	15.05	16.84
φ / radians	-1.759	-1.659
ω / radians	1.045	1.035
κ / radians	-2.973	-3.048

Table 5.10: Recovered camera extrinsic parameters for test image img_1707.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364242.09	364239.85
Y / meters	143468.03	143469.47
Z / meters	14.56	15.79
φ / radians	-2.989	-2.919
ω / radians	-1.289	-1.308
κ / radians	1.734	1.794

Table 5.11: Recovered camera extrinsic parameters for test image pa091640.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364358.46	364358.99
Y / meters	143448.86	143448.25
Z / meters	14.72	15.90
φ / radians	-1.741	-1.716
ω / radians	-0.046	-0.031
κ / radians	3.125	3.129

Table 5.12: Recovered camera extrinsic parameters for test image pa091642.

Extrinsic Parameters	Pose Estimated	Hybrid Space Resection
X / meters	364335.92	364337.25
Y / meters	143448.73	143447.32
Z / meters	15.75	15.83
φ / radians	-1.691	-1.688
ω / radians	-0.605	-0.604
κ / radians	3.075	3.059

Table 5.13: Euclidean distance between pose estimated and hybrid space resection recovered camera positions acquired by various cameras.

Image name*	Euclidean distance between hybrid recovered position and pose estimated position / meters
dsc_5403	1.07
dsc_7967	1.16
img_1707	2.93
img_7089	1.94
pa091642	1.43
pa091640	1.94

* dsc_XXXX denotes image acquired by Nikon D50, img_XXXX denotes image acquired by Canon S5 Pro, paXXXXXX denotes image acquired by Olympus E-420

Figures 5.1 – 5.3 shows OpenGL pose of the building model rendered by the Graffiti program. Using the current estimate pose of the camera, the corresponding AOI polygon for each facade was generated and overlaid on the terrestrial photograph to evaluate the accuracy of the recovered parameters. We observed that the AOI polygon lines align themselves with the edges of the building façade on the terrestrial photograph, denoting a very good estimate of the extrinsic parameters from the pose of the building models used.

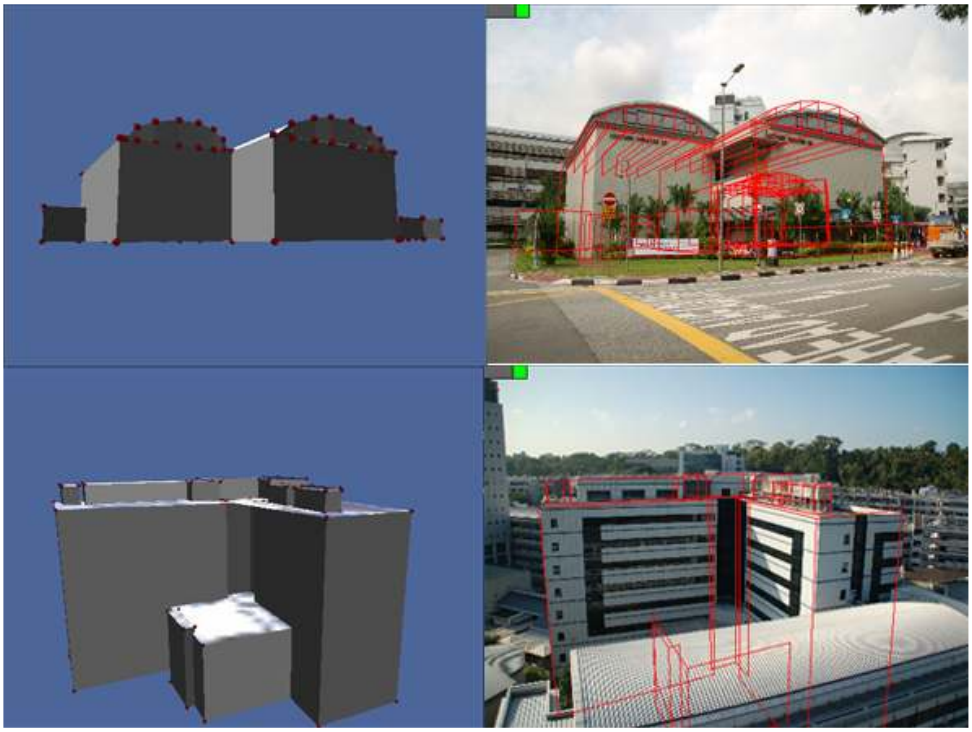

Figure 5.1: Selected pose for building model (left) and corresponding façade AOI polygon (right) on terrestrial photographs for dsc_5403 (top) and dsc_7976 (bottom).

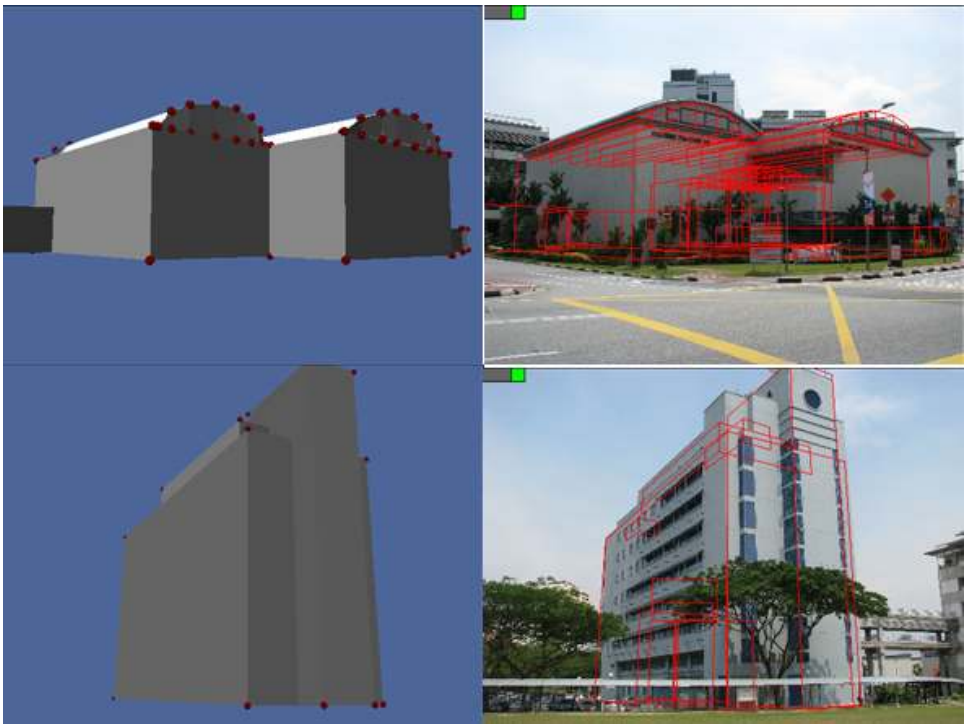

Figure 5.2: Selected pose for building model (left) and corresponding façade AOI polygon (right) on terrestrial photographs for img_1707 (top) and img_7089 (bottom).

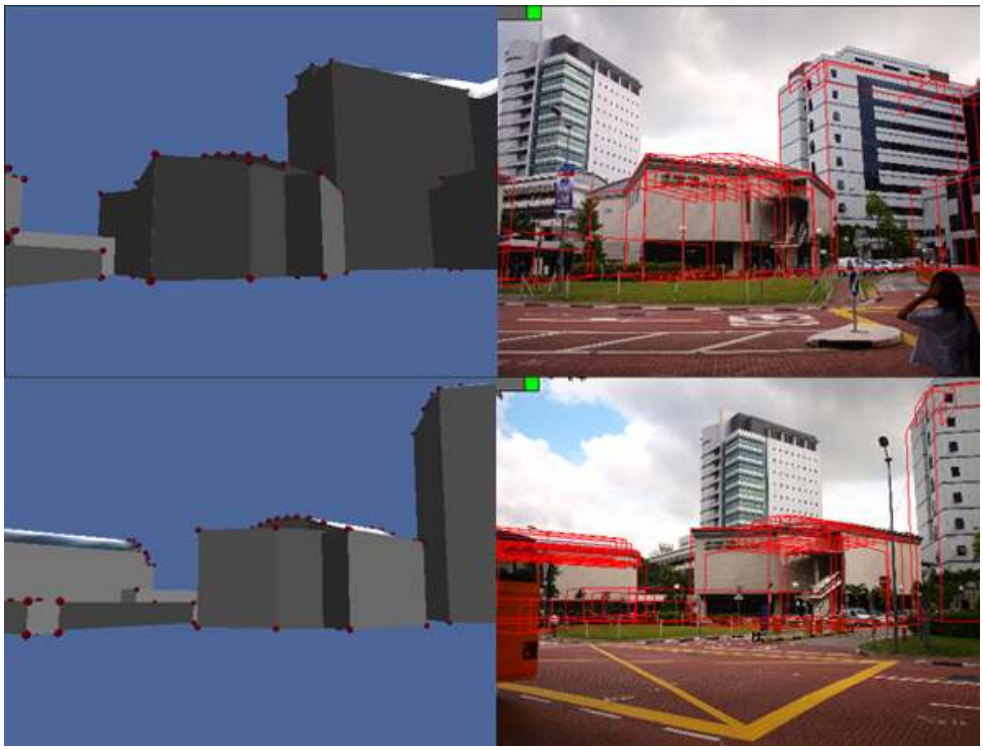

Figure 5.3: Selected pose for building model (left) and corresponding façade AOI polygon (right) on terrestrial photographs for pa091642 (top) and pa091640 (bottom).

5.1.3 Discussion for Camera Extrinsic Parameters Recovery Methods

In the previous two sections, we attempt to evaluate the recovery performance and validate the accuracy of the recovered extrinsic parameters using both methods. Firstly, hybrid space resection has been shown to recover the extrinsic parameters accurately. Furthermore, it is robust when data points have been contaminated by noise. This is because least squares adjustment technique which was integrated into the algorithm is able to reduce the effect of outliers. However, measurement errors of 0.1 mm are unlikely on the image as the actual offset in terms of pixels is relatively large. Depending on the type of digital camera and its respective sensor size, the pixel error could range anything between 12.6 pixels to 44.4 pixels for an error of 0.1mm. Conversely, measurement errors of 1.00 meters for ground control points are more likely to occur as the 3-D building models we used were extracted from stereo-pair

IKONOS satellite images which have a pixel resolution of 1.0 meter at nadir and accuracy of ± 1.5 meters (Wang *et al*., 2005).

Secondly, interactive pose estimation has been shown to provide a good estimate of the extrinsic parameters without performing any ground control points and image control points matching. Pose estimation can be used to recover the extrinsic parameters of the camera in a quick and efficient manner when hybrid space resection fails to work in event of gross errors occurring in the prior phase of building 3-D reconstruction or when the geo-referencing information of 3-D building model is absent.

5.2 Evaluation of Quality of Extracted Façade Textures

In a virtual city environment, the building models need to be represented in a photorealistic and convincing manner. As discussed earlier in Chapter 4, terrestrial photographs taken from multiple angles and different physical conditions can be used to texture the building models. A mosaic composition of two or more images can be used as well if the wall façade is too width. The quality of the extracted textures can be evaluated in terms of the geometric accuracy and colour continuity.

5.2.1 Geometric Accuracy of Extracted Façade Textures

The geometric accuracy of the extracted façade textures can be qualitatively accessed by comparing the extracted textures taken from both the frontal normal view as well as the oblique view. The general assumption is that the façade surfaces to be textured are planar in general and do not have any additional features protruding from the surface. This is because in order to perform geometric correction accurately on such features, we will need the precise vertex information belonging to them. Such information is lacking in the standard 3-D building model data sets. Terrestrial photographs are taken from both normal and oblique view and processed using the IRPTM method as described in Chapter 4. The textures are automatically extracted and geometrically corrected based on the location and orientation parameters of the photographs.

Figure 5.4 shows the top view of the test building model which was used in this experiment. The building model was extracted using a pair of 1 meter resolution satellite imagery and the terrestrial photographs were taken using commercially available cameras. The red box denotes the test façade for which we will perform texture mapping on. The positions and orientations of both the normal and oblique

photographs are annotated on Figure 5.4 as well. Figure 5.5 shows the textures extracted from both the normal and oblique view images for the test façade. Geometric correction was performed on both textures and this was dependent on the position and orientation of the camera when the photographs were taken. With the exception of different objects situated in front of the test façade, there was little observable difference between the extracted textures taken from both viewing perspectives.

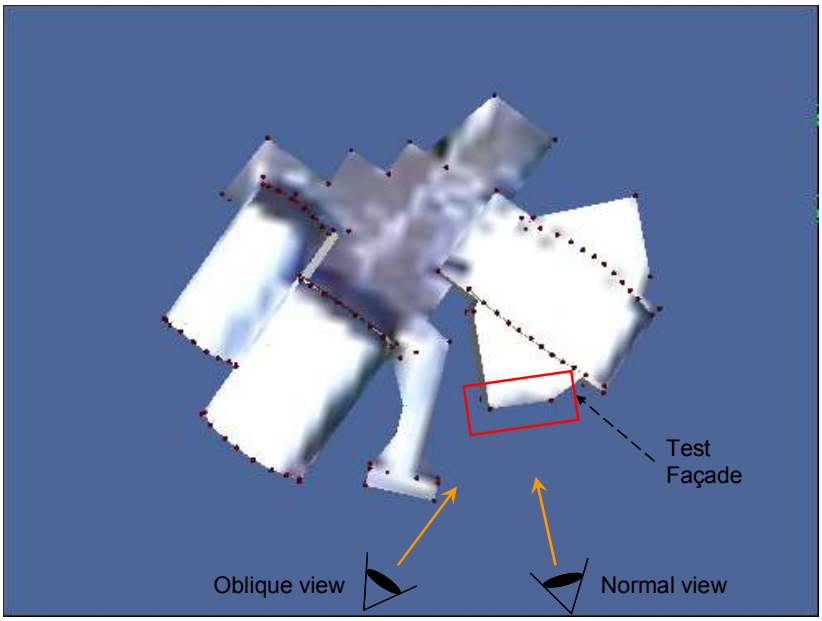

Figure 5.4: Top view of 1st building model.

Geometrically corrected normal view texture

Geometrically corrected oblique view texture

Figure 5.5: Geometrically corrected façade textures.

Figure 5.6 shows the top view of another test building used in our evaluation. The setup was similar to the first test building with 2 photographs taken in both the normal and oblique view. Geometric correction was performed on both extracted textures and the observations were similar to the first test. There was no difference between the extracted textures except for objects which were not part of the façade.

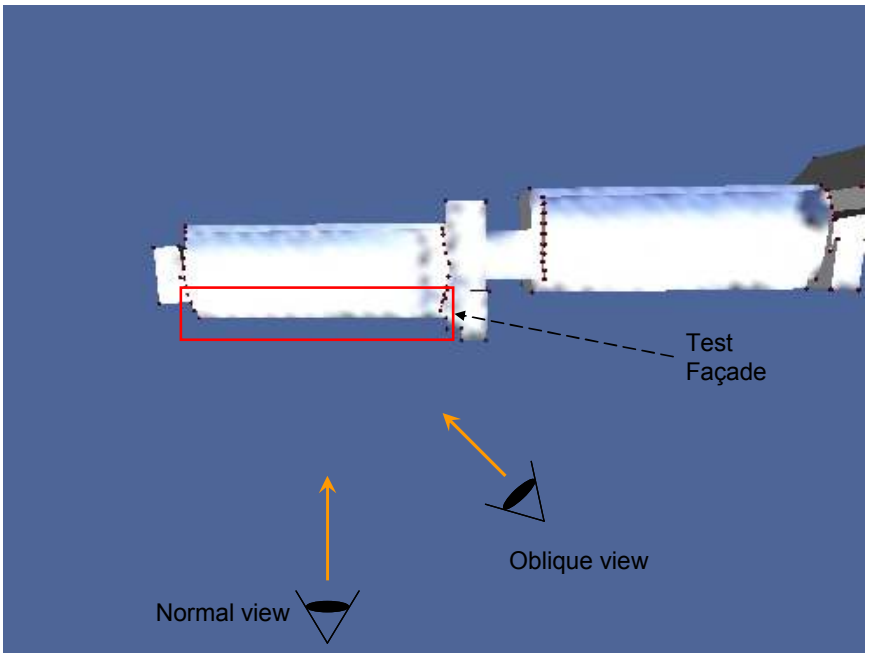

Figure 5.6: Top view of 2nd test building.

Geometrically corrected normal view texture

Geometrically corrected oblique view texture

Figure 5.7: Geometrically corrected façade textures.

5.2.2 Colour Continuity of Façade Texture Mosaic

After the geometrically corrected textures have been extracted from various terrestrial photographs, we proceed to perform colour balancing. It is possible that the photographs were taken under different illumination conditions, hence colour balancing needs to be done to ensure a smooth continuous flow between adjacent façade textures. In our evaluation setup, a test 3-D building model was textured using different photographs taken using different cameras at different times of the day. As shown in Figure 5.8, the images were acquired during different times of the day to obtain varying illumination conditions on the building façades. The date and time information can be found in Table 5.14. Façade textures were extracted and geometrically corrected. They were subsequently placed together, forming a mosaic composition of façade textures as shown in Figure 5.9.

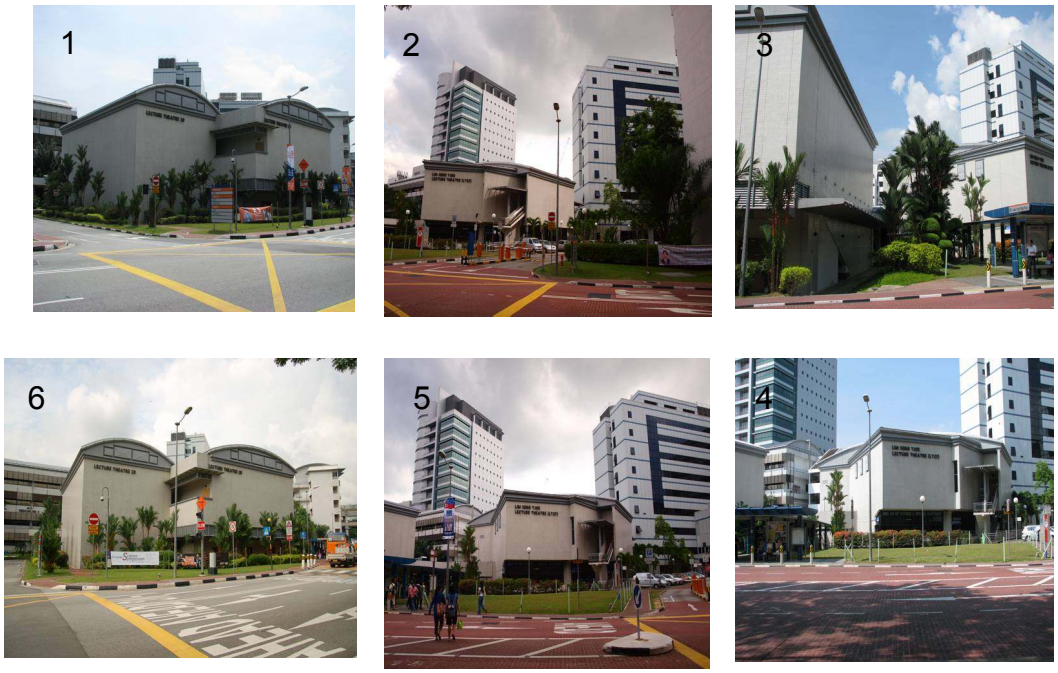

Figure 5.8: Different terrestrial photographs used for texturing.

Table 5.14: Additional information regarding terrestrial photographs.

s/n	Camera Type	Date	Time
1	Canon	28/02/2008	1455
2	Olympus	09/10/2008	1528
3	Canon	09/05/2008	1051
4	Canon	09/05/2008	1050
5	Olympus	09/10/2008	1529
6	Nikon	17/10/2008	0956

Non colour balanced façade mosaic Colour balanced façade mosaic

Figure 5.9: Façade texture mosaic before (left) and after (right) colour balancing.

It was observed that the colour balanced façade texture mosaic as shown in Figure 5.9 (right) showed a more continuous texture flow between adjacent façades as compared to the non colour balanced variant (left). We performed a quantitative analysis on the histogram of the façade mosaic as shown in Figures 5.10 – 5.12. It was observed that the standard deviations of the colour intensities were smaller compared to the non colour balanced mosaic. This means that the spread of pixel intensity values were reduced, making the texture mosaic more uniform and gradual in transition between adjacent façades. Figure 5.13 shows the building model textured with the colour balanced façade textures. It was observed that building textures were uniform even when they were acquired by different cameras and conditions.

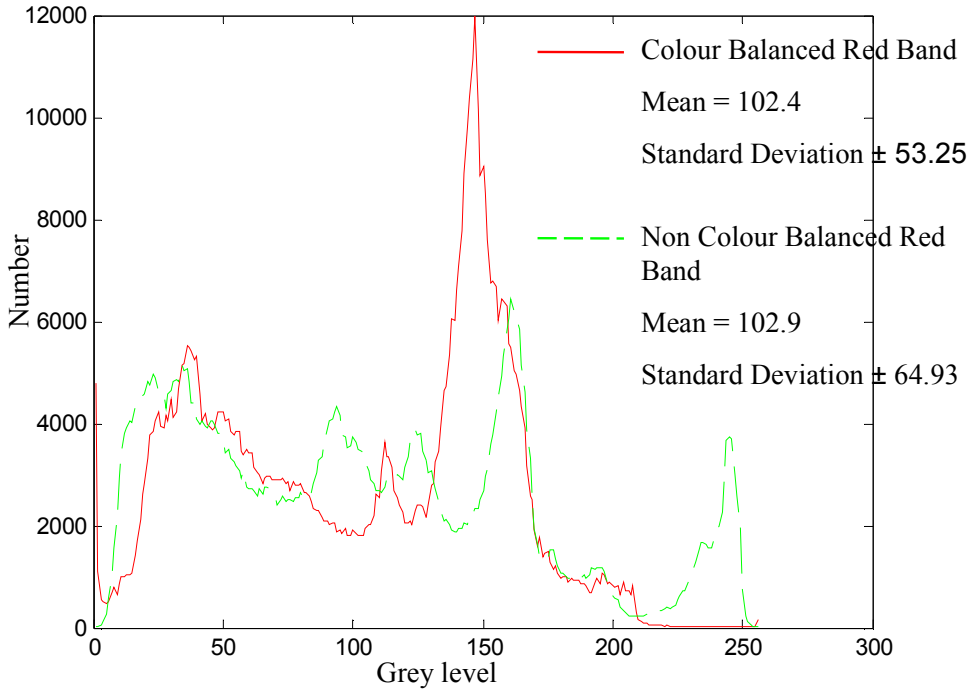

Figure 5.10: Red band histogram of colour balanced (red) and non colour balanced (green) façade mosaic.

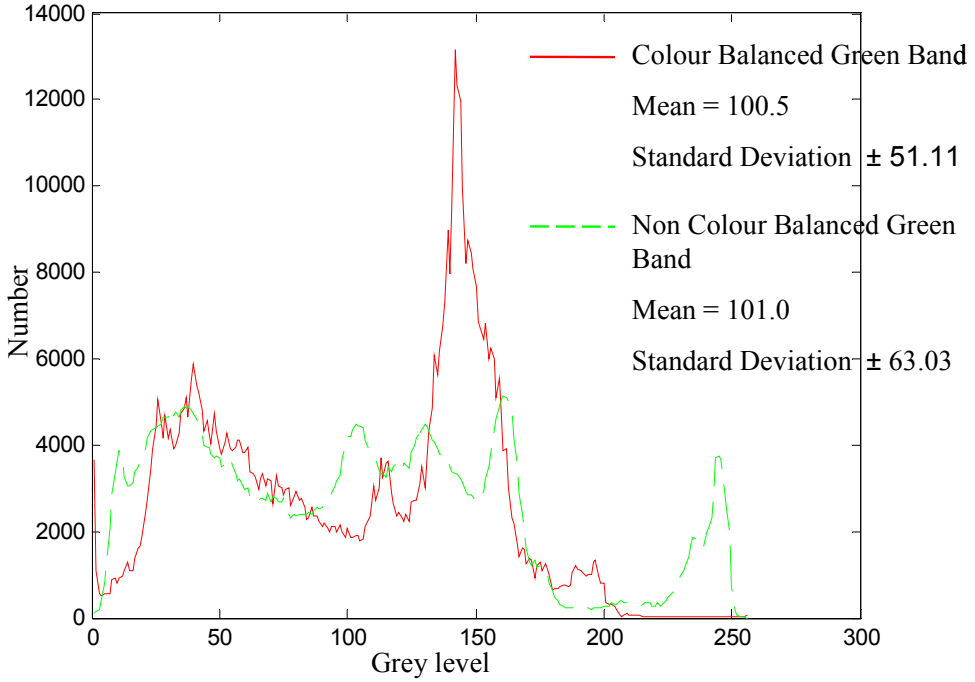

Figure 5.11: Green band histogram of colour balanced (red) and non colour balanced (green) façade mosaic.

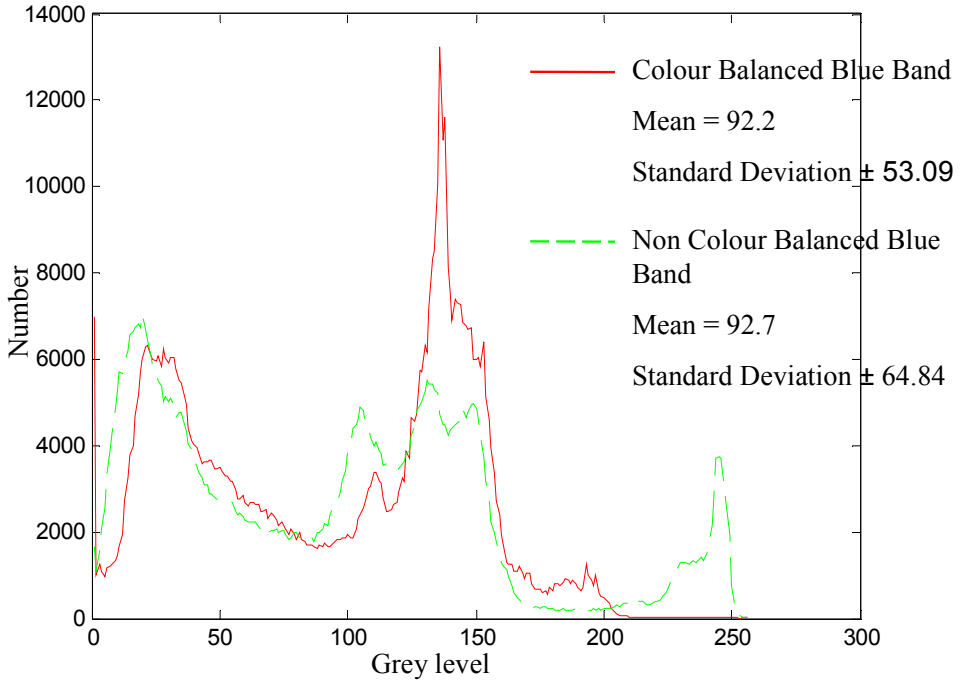

Figure 5.12: Blue band histogram of colour balanced (red) and non colour balanced (green) façade mosaic.

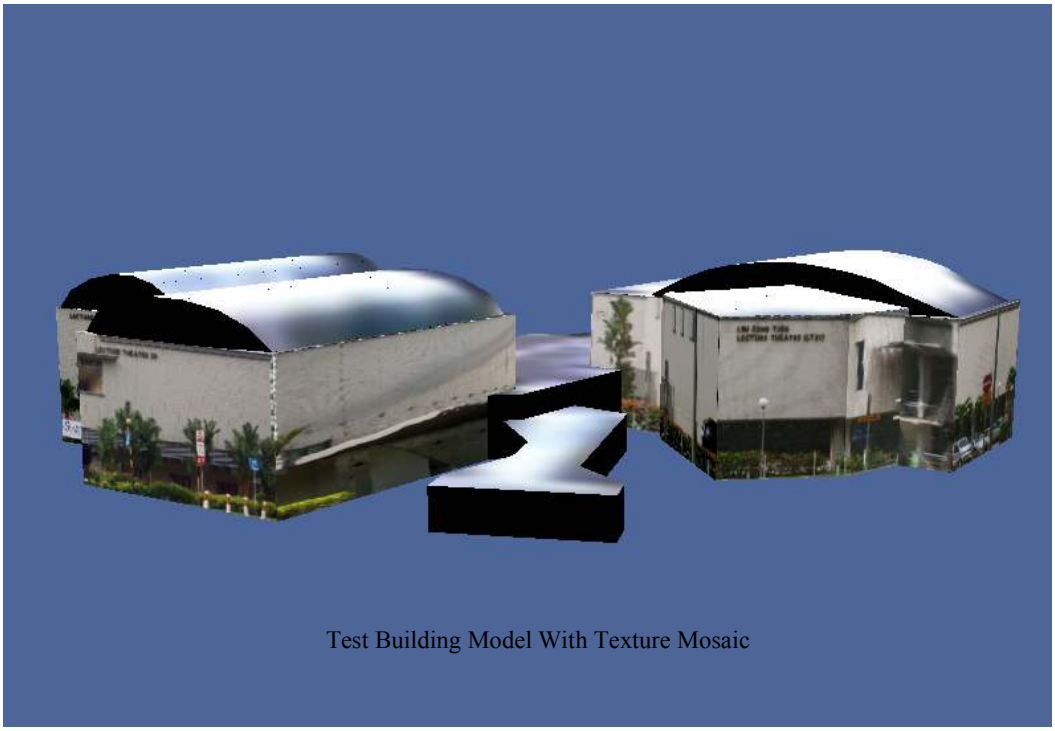

Figure 5.13: Test building model textured with colour balanced texture mosaic.

5.2.3 Visual Quality of Mosaic Composition

A very wide building façade was chosen to be textured such that it was only possible to capture a complete image of the façade from either a single shot from a very narrow oblique view (position B) or by two partial near oblique view shots (position A1 and A2) as shown in Figures 5.14 and 5.15. Correspondence between images taken from positions A1 and A2 was established by taking more than four control points on each respective image. A mosaic composition was created using the method described in Section 4.5 and shown in Figure 5.16. The required texture was subsequently extracted automatically using the IRPTM method and compared to the texture extracted from the very narrow oblique view image.

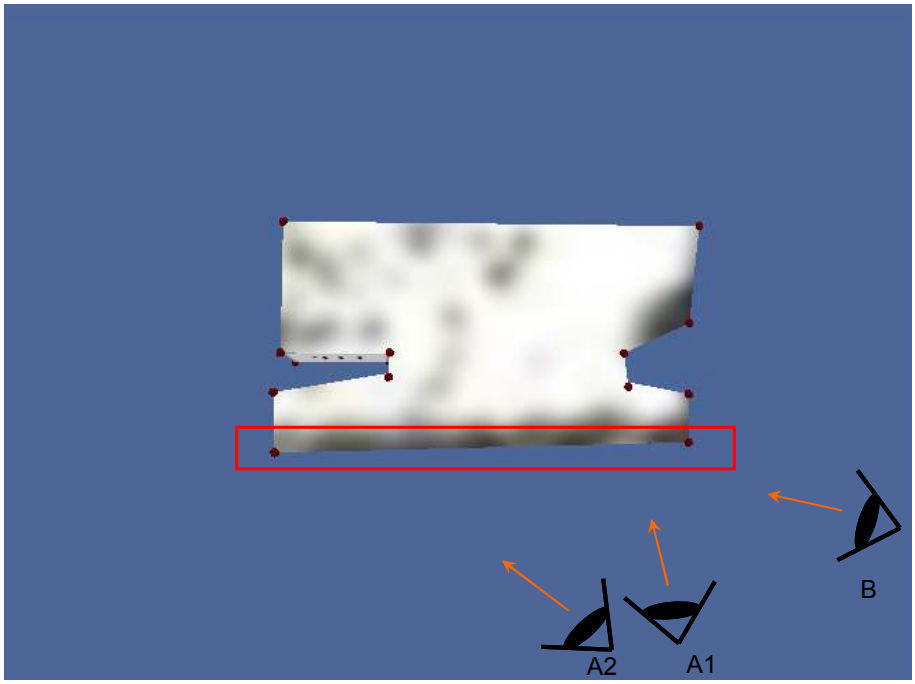

Figure 5.14: Position of cameras used for texturing a wide building façade.

Figure 5.15: Terrestrial photographs taken from the various view points.

Figure 5.16: Resultant mosaic composite from A1 and A2

It was observed in Figure 5.17 that the texture extracted from the mosaic composite produced a more visually accurate rendition of the actual façade. This was due to the perspective foreshortening effect which was not taken into account in the geometric correction. As depth information of the façade was not incorporated, this resulted n the slightly skewed image observed right image of Figure 5.17.

Figure 5.17: Extracted façade from mosaic composite (left) and narrow oblique image (right).

5.2.4 Discussion for Quality of Extracted Façade Textures

In the previous two sections, we attempt to evaluate the quality of the extracted façade textures. Firstly, it is important for the extracted façade textures to retain the geometric accuracy of the original façade features in order to texture the building model in a realistic and convincing manner. In order to reduce the number of photographs required to texture the building, oblique view photographs of the building are taken. The extracted textures are inevitably distorted due to the view angle but the distortion can be effectively removed via the geometric correction method as described in chapter 4. We have shown in the previous section that after geometric correction, there is no observable difference between the extracted texture from an oblique view image and a normal view image.

Secondly, we need to ensure a smooth and gradual transition between adjacent façade textures as the photographs were taken under different illumination and physical conditions. This results in differences in colour, brightness and contrast between adjacent façade textures. We have shown that the differences in brightness and contrast between adjacent façade textures can be reduced by performing colour balancing and this will improve the overall visual effect of the composite building texture.

Finally, it is not always possible to capture the entire façade using just a single image due to physical limitations of the camera lens and sensor. However, we can overcome this limitation by creating a mosaic composite comprising of several images of the façade and successfully extract an accurate texture representation of the building façade.

Chapter 6

Conclusions

The main objective was to perform texture mapping using terrestrial photographs without the help of external direct referencing sensors on standard 3-D building models in a quick and efficient manner. This was done in order to improve the photorealistic effect of the building models. We have achieved that through the IRPTM implementation in Graffiti which was presented in this thesis.

Firstly, it was shown that we were able to retrieve the camera's extrinsic parameters efficiently without the help of external sensors. This was done using two techniques, namely hybrid space resection and pose estimation. The accuracy of the extrinsic parameters recovered by both methods was verified using pre-existing data sets and yielded satisfactory results. They were compared to the traditional non-linear space resection method and the relative performance was superior. Highly detailed building models were not required as pose manipulation enabled the user to fit the rendered scene to match the image. The reverse projection of the façade AOI polygons onto the terrestrial image was done in a novel way to confirm the extrinsic parameters of the camera.

Secondly, in order to perform texture extraction in a quick and efficient manner, we can reduce the number of photographs needed by taking oblique view images possibly under different conditions and using different imaging equipment. As a result, distortions which were highly dependent on the extrinsic parameters were introduced but shown to be removed effectively using the described geometric correction method. Most illumination induced radiometric differences between

adjacent façade textures were reduced as well after adopting a colour balancing algorithm. The entire process of texturing a building model can be done in a short time. The test building used in our evaluation as shown previously in Figure 5.13 was fully textured in less than 15 minutes by a skilled operator. This is only a fraction of the time taken by other methods previously used.

In conclusion, by integrating the extrinsic parameters recovery, geometric correction and color balancing methods, we are able to enhance the capabilities of the IRPTM method, making it superior when compared to FTM and the existing RPTM techniques.

6.1 Further Work

By using the IRPTM implementation in the Graffiti program which we have developed for texture extraction, it opens several avenues for further work which we can look into.

6.1.1 Façade Enhancement

Texturing of the building models is just the first step to enhance the realism of the city model in a mixed reality environment. Several approaches can be taken to improve the detail of the building models based on the current data set which we have at hand. Firstly, due to the multiple photographs taken of the target building, texture information redundancy has been built up. Building façades are often blocked by various natural or man-made objects and this poses a problem during image acquisition as they are captured in the photograph as well. With the redundancy in texture data, it is theoretically possible to reduce the amount of occlusion on the façade texture.

Secondly, after we have located the extrinsic parameters of the terrestrial photographs, it will be a logical step to proceed with the 3-D reconstruction of the façade as a stereo-pair setup is obtained.

6.1.2 Automated Pose Estimation

Automated pose recovery of a photograph is still a very challenging subject due to the number of variables involved. The pose of a camera taking the photograph has 6 degrees of freedom and this will result in a very large search space to recover the parameters automatically. However, we can reduce the complexity of the problem after the texturing of the building models are complete. By accurately recreating the city in a virtual environment with the textured building models and terrain elevation, there will be more information available for correlation matching between the rendered scene and actual photograph. There is great potential to increase the robustness of the automated pose estimation algorithm in this approach.

References

[1] Aurigi. A. and De Cindio, F. Augmented Urban Spaces, Articulating the Physical and Electronic City. England: Ashgate Publishing Limited. 2008.

[2] Ballard, D.H. and Brown, C.M. Computer Vision. New Jersey: Prentice-Hall. 1982.

[3] Baltsavias, E., Grun, A. and van Gool, L. Automatic Extraction of Man-made Objects from Aerial and Space Images (III). Lisse: A.A Balkema. 2001.

[4] Beck, M. "Real-time visualization of big 3D city models". International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV-5/W10. CD-ROM. 2003.

[5] Canny, J.F. "A computational approach to edge detection". IEEE PAMI, pp 679-698. 1986.

[6] Coorg, S. and Teller, S. "Spherical mosaics with quaternions and dense correlation". International Journal of Computer Vision 37(3), pp 259-273. 2000.

[7] Coorg, S. and Teller, S. "Extracting textured vertical facades from controlled close-range imagery". In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp 625-632.

[8] Coors, V., Huch, T. and Kretschmer, U. "Matching buildings: pose estimation in an urban environment". In Proc. IEEE and ACM International Symposium on Augmented Reality, 2000, Munich, Germany, pp 89-92.

[9] Debevec, P.E., Camillo, J.T. and Malik, J. “Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach”. In Proc. International Conference on Computer Graphics and Interactive Techniques, 1996, pp 11-20.

[10] Gonzales, R.C. and Woods, R.E. Digital Image Processing 3rd Edition. pp 94-100, New York: Prentice Hall Inc. 2007.

[11] Google Earth™ mapping service, http://www.earth.google.com.

[12] Grodecki, J. “IKONOS stereo feature extraction - RPC Approach”. In Proc. Of ASPRS 2001 Conference, 2001, St. Louis, USA, CD-ROM.

[13] Grodecki, J. and Dial, G. “IKONOS geometric accuracy”. In Proc Of ISPRS Workshop on High Resolution Mapping from Space, 2001, Hanover Germany, pp 77-86.

[14] Haala, N. and Bohm, J. “A multi-sensor system for positioning in urban environments”. ISPRS Journal of Photogrammetry & Remote Sensing, Vol. 58, pp 31-42, 2003.

[15] Hoff, B. and Azuma, R. “Auto-calibration of an electronic compass in an outdoor augmented reality system”. In Proc. IEEE and ACM International Symposium on Augmented Reality, 2000, Munich, Germany, pp 159-164.

[16] Huang, X., Kwoh, L.K., Yuan, B. and Tan, Y.K.A. “An efficient platform for 3D city model visualization”. In Proc. IEEE International Geoscience and Remote Sensing Symposium, 2006, Denver Colorado, USA, pp 297-300.

[17] Kass, M., Witkin, A.and Terzopoulos, D. Snakes: “Active contour models”. International Journal of Computer Vision, 1987, pp321-331.

[18] Kumar, R., Sawhney, H.S., Guo, Y., Hsu, S. and Samarasekera, S. "3D manipulation of motion imagery". In Proc. International Conference on Image Processing, 2000, Vancouver, BC, Canada, pp 17-20 Vol. 1.

[19] Kwoh, L. K. "Processing of stereo Spot imagery for coordinate determination". Masters Thesis, University of New South Wales, 1988.

[20] Laycock, R.G. and Day, A.M. "Automatic techniques for texture mapping in virtual urban environments". In Proc. Of the Computer Graphics International (CGI 04) IEEE, 2004, pp 586 - 589.

[21] McGlone, C. (ed). Manual of Photogrammetry, 5th Edition USA: American Society for Photogrammetry, 2004.

[22] Milgram, P. and Kishino, F. "A taxonomy of mixed reality visual displays". ICICE Transactions on Information Systems, Vol. E77-D, No. 12 December 1994.

[23] Mikhail, E.M., Bethel, J.S. and McGlone, J.C. Introduction to modern photogrammetry. USA: John Wiley & Sons, 2001.

[24] Nicolas, H. "New methods for dynamic mosaicking". IEEE Transactions on Image Processing, 10(8) pp 1239-1251, 2001.

[25] Ortin, D. and Remondino, F. "Occlusion-free image generation for realistic texture mapping". In Proc. ISPRS Working Group V/4, 2005.

[26] Scharl, A. "Towards the geospatial web: Media platforms for managing geotagged knowledge repositories". The Geospatial Web - How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society, pp 3-14, London: Springer, 2007.

[27] Tsai, F. and Lin, H.C. “Polygon-based texture mapping for cyber city 3D building models”. International Journal of Geographical Information Science, 21:9, pp 965-981, 2007.

[28] Wang, J., Di, K. and Li, R. “Evaluation and improvement of geo-positioning accuracy of IKONOS stereo imagery”. Journal of Surveying Engineering, Vol. 131(2), pp 35-42, 2005.

[29] Wolf, P.R. and Dewitt, B.A. Elements of Photogrammetry with Applications in GIS, 3rd Edition. New York: McGraw-Hill. 2000.

[30] Wright, R.S. and Lipchak, B. OpenGL Super Bible, 3rd Edition. USA: Sams Publishing. 2005.

[31] Zeng, Z. and Wang, X. “A general solution of a closed-form space resection”. Photogrammetric Engineering & Remote Sensing, Vol. 58(3), pp 327-338, 1992.

[32] Zhang, Z., Kang, Z. “The rendering of building texture from land-based video”. In Proc of XXth ISPRS Congress Commission 3, 2004, Istanbul, Turkey, CD-ROM.

Appendix A

List of Test Data Sets

Set 1

This test data set was obtained from Wolf and Dewitt (2000).

	Actual Position
X / meters	1027.86
Y / meters	1044.12
Z / meters	648.20
φ / radian	-0.00717
ω / radian	0.02112
κ / radian	1.7942

Set 2

This test data set was obtained from Kwoh (1988).

	Actual Position
X / meters	9157.27
Y / meters	17864.07
Z / meters	1562.13
φ / radian	0.01429
ω / radian	0.025359
κ / radian	-1.79624

Set 3

This test data was obtained from Kwoh (1988).

	Actual Position
X / meters	8935.93
Y / meters	17032.27
Z / meters	1581.52
φ / radian	-0.001605
ω / radian	0.039915
κ / radian	-1.79768

Appendix B

List of Publications

[1] Tan, Y. K. A., Kwoh, L. K. and Ong, S. H., "Large scale texture mapping of building facades", In Proc. of XXI Congress, International Society for Photogrammetry and Remote Sensing, 2008, Beijing, China, pp 687-692.

[2] Tan, Y. K. A., Kwoh, L. K. and Ong, S. H., "Texture mapping of 3-D building models using pose estimation of digital photographs", In Proc. of 29th Asian Conference on Remote Sensing (ACRS2008), 2008, Colombo, Sri Lanka.