Novel Approach to Epipolar Resampling of HRSI and Satellite Stereo Imagery-based Georeferencing of Aerial Images

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jaehong Oh, M.S., B.S.

Graduate Program in Geodetic Science and Surveying

The Ohio State University

2011

Dissertation Committee:

Dorota A. Grejner-Brzezinska, Adviser

Alper Yilmaz

Ralph R. von Frese

Charles K. Toth
ABSTRACT

The last decade saw major breakthroughs in geospatial image technology in both the spaceborne and in airborne domains. The commercial use of high-resolution satellites was enabled by the Presidential Decision Directive 23 issued by the Clinton Administration in 1994. At the same time, high-performance large-format digital cameras have been replacing traditional large-format aerial film cameras. Both technologies are completely digital and provide high image quality as well as efficient data acquisition capability. However, there are still significant technical challenges in exploiting these images, and thus in achieving high-accuracy 3D mapping performance.

An ongoing challenge in high-resolution satellite image (HRSI) processing has been the establishment of accurate epipolar geometry over the entire HRSI image area, which is a pivotal step of any stereo-image processing, including 3D topographic mapping. However, the pushbroom camera which is used by most high-resolution satellites does not produce straight epipolar lines. Furthermore, in contrast to the well-known frame cameras, the epipolar curve pair does not exist for the entire stereo image area. These properties make it difficult to establish the epipolar geometry of a pushbroom camera, resulting in a limited accuracy of the epipolar image resampling. In this study, a new method of epipolar curve pair determination and epipolar image resampling of
spaceborne pushbroom imagery, based on the popular RPC (Rational Polynomial Coefficients), is proposed. First, the new method determines the epipolar curve pairs over the stereo HRSI for establishing the epipolar geometry. Next, accurate epipolar image resampling is achieved by arranging the determined epipolar curve pairs over the stereo images to satisfy the epipolar resampled image condition. In this research, the characteristics of epipolar curves in the image were investigated, whose stereo sensor model is RPC, and the proposed approach was tested and validated. The experimental results showed that the new method could not only successfully establish the epipolar curves over the entire stereo HRSI, which enables the accurate epipolar image resampling, but also showed superior results with better accuracy than the conventional method, due mainly to a simpler model that avoids complex geometry of the pushbroom camera.

In airborne imaging systems, one challenge is that alternative automated georeferencing is required to serve as a backup for direct georeferencing in case when GPS/INS-based (Global Positioning System/Inertial Navigation System) georeferencing is not feasible. This is particularly important when the GPS accuracy or signal availability is not assured, which can jeopardize the mapping accuracy; note that GPS is a line-of-sight system, and subject to interference and jamming. Moreover, there are areas that completely lack the ground-based GPS infrastructure, which is often required to maintain mapping accuracy by differential GPS methods. Therefore, in this study, the potential of stereo HRSI as ground control information for georeferencing of aerial imagery is investigated. High spatial and temporal resolutions of HRSI, their large swath width, and
high positional accuracy, for inaccessible areas, motivated this research. Note that if a monoscopic HRSI is used as a ground control, apparent leaning of the on-ground objects in HRSI introduce positional errors to the horizontal ground coordinates. Consequently, external ground elevation data are needed to provide ground height information. In contrast, the use of a stereo HRSI can provide the 3D ground control information by 3D ground restitution, and avoids the aforementioned problem. To implement the proposed approach, a robust and efficient image matching scheme was developed to handle potential image differences between aerial images and stereo HRSI. The method is based on a multi-scale image matching approach, utilizing combination of SIFT (Scale-Invariant Feature Transform) and RANSAC (Random Sample Consensus) methods. In the georeferencing step, the bundle adjustment with outlier removal based on Baarda’s data snooping is utilized. The proposed method was tested for a strip of digital aerial images and a set of IKONOS stereo images. Even though a large seasonal difference between the image sources, the matching scheme was able to successfully extract 3D ground control information from the stereo IKONOS images by mitigating the outlier matching points. Also, georeferencing of aerial images that lack distinct image features, such as forest area, was also possible by utilizing tie points to adjacent aerial images. One-meter horizontal and 1.5 meter vertical ground positional accuracy was achieved, which is rather good, considering 1-meter resolution of IKONOS images. The fully-automated approach has potential not only for aerial image georeferencing practically all
over the world, including inaccessible areas, but also provides a good basis for image-based aerial navigation under GPS denied-conditions.
DEDICATION

I dedicate this dissertation to my wife Youngsook, and my precious daughter Eunchae. Youngsook sacrificed her life to support my studies. She was always on my side and prayed for me. The birth of Eunchae was the greatest moment of my life and she gave me a new appreciation for the meaning and importance of family. I love you.
ACKNOWLEDGMENTS

My thanks and appreciation to Prof. Dorota A. Grejner-Brzezinska for her support as my advisor throughout the time it took me to complete this research and write the dissertation. Also, I wish to express deepest appreciation to Dr. Charles K. Toth for his professional guidance as my co-advisor throughout the whole graduate study. I thank the members of my dissertation committee, Prof. Alper Yilmaz and Prof. Ralph R. von Frese, for their contribution and their good-natured support.

I am very grateful to Prof. Changno Lee of Seoul National University of Technology, who gave me many valuable ideas and advice. I need to express my gratitude and deep appreciation to Prof. Hyosung Lee of Suncheon National University and Prof. Dongyeop Han of Chonnam National University for their supporting me during my graduate study.

I must acknowledge as well to Dr. Karla Edwards and Omar Mora of SPIN lab for the help to improve this dissertation and also thank the SPIN lab colleagues, Dr. Shahram Moafipoor, Xiankun Wang, Jihye Park, Yuksel Boz, Hui Ju, Amir Osman, Siavash HosseinyAlamdary, and Leilei Li. Thanks also to friends who studied together at OSU, Youngjin Lee, Jinwoo Kim, Yunjae Choung, Raesung Kim, Hyojin Yang.

I am also grateful to brothers and sisters in Korean Church of Columbus.

Finally, a special thank to my parent for their sacrifice and love.
VITA

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 1976</td>
<td>Born - Masan, Korea (ROK)</td>
</tr>
<tr>
<td>February 1999</td>
<td>B.S., Urban Engineering, Seoul National University, Seoul, Korea (ROK)</td>
</tr>
<tr>
<td>February 2001</td>
<td>M.S., Survey and Remote Sensing, Seoul National University, Seoul, Korea (ROK)</td>
</tr>
<tr>
<td>March 2001 – April 2004</td>
<td>Researcher, Space Imaging Asia, Yongin, Korea (ROK)</td>
</tr>
<tr>
<td>May 2004 – August 2007</td>
<td>Researcher, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea (ROK)</td>
</tr>
<tr>
<td>September 2007 – June 2010</td>
<td>Graduate Research Associate, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>April 2010</td>
<td>2010 the American Society for Photogrammetry &amp; Remote Sensing (ASPRS) Intergraph scholarship and the annual BAE Systems award</td>
</tr>
<tr>
<td>June 2010</td>
<td>M.S., Geodetic Science, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>May 2011</td>
<td>2011 ASPRS Robert E. Altenhofen Memorial</td>
</tr>
</tbody>
</table>
Scholarship

July 2010 - Present
Presidential Fellow, the Graduate School, The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Geodetic Science and Surveying

Photogrammetry and remote sensing
TABLE OF CONTENTS

ABSTRACT ........................................................................................................... ii
DEDICATION ......................................................................................................... v
ACKNOWLEDGMENTS ......................................................................................... vi
VITA ...................................................................................................................... vii
TABLE OF CONTENTS .......................................................................................... ix
LIST OF TABLES .................................................................................................. xii
LIST OF FIGURES ................................................................................................ xiv
1 INTRODUCTION .................................................................................................. 1
  1.1 Epipolar resampling of HRSI ................................................................. 3
  1.2 Satellite Stereo Imagery-based Georeferencing of Aerial Images .......... 9
  1.3 Dissertation outline ................................................................................. 14
2 OVERVIEW OF EPIPOLAR GEOMETRY AND RESAMPLING .................. 16
  2.1 Frame camera ......................................................................................... 17
    2.1.1 Epipolar line equation ................................................................. 22
    2.1.2 Epipolar image resampling ......................................................... 30
  2.2 Pushbroom camera ................................................................................. 38
    2.2.1 Rigorous model .............................................................................. 40
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2</td>
<td>2D Affine model</td>
<td>46</td>
</tr>
<tr>
<td>2.2.3</td>
<td>RPC- Replacement Sensor Model</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>EPIPOLAR IMAGE RESAMPLING OF SATELLITE IMAGES</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Properties of the epipolar curves from RPC</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Epipolar curve generation and resampling</td>
<td>66</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Piecewise epipolar curve generation</td>
<td>67</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Epipolar resampling</td>
<td>69</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Generation of the RPC for the epipolar resampled imagery</td>
<td>74</td>
</tr>
<tr>
<td>3.4</td>
<td>Test and evaluation</td>
<td>76</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Test data specification and preparation</td>
<td>76</td>
</tr>
<tr>
<td>3.4.2</td>
<td>The piecewise epipolar curve generation</td>
<td>80</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Epipolar resampling and accuracy analysis</td>
<td>85</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>AERIAL IMAGES GEOREFERENCING USING STEREO HRSI</td>
<td>96</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>Image matching between aerial and satellite images</td>
<td>103</td>
</tr>
<tr>
<td>4.2.1</td>
<td>SIFT matching</td>
<td>103</td>
</tr>
<tr>
<td>4.2.2</td>
<td>SIFT performance analysis</td>
<td>110</td>
</tr>
<tr>
<td>4.2.3</td>
<td>RANSAC</td>
<td>124</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Multi-scale approach</td>
<td>130</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1 Currently operating high-resolution satellite image (HRSI) systems.......... 4
Table 1.2 Aerial digital cameras. ................................................................................. 10
Table 3.1 Straight line fitting orthogonal direction residual of local epipolar curves. ..... 60
Table 3.2 Test data specification. ..................................................................................... 76
Table 3.3 RPC residuals in the left and right image before and after the RPC update using GCP.................................................................................................................. 79
Table 3.4 Residuals of each polynomial epipolar resampling equation. ....................... 88
Table 3.5 y-parallax for tie points by the proposed method. ........................................... 89
Table 3.6 y-parallax for 16 measured tie points estimated by the parallel projection. ..... 90
Table 3.7 Ground restitution RMS for the proposed method. ....................................... 90
Table 4.1 Some examples of RANSAC threshold with probability 95%....................... 128
Table 4.2 FFT based normalized cross-correlation matching performance. ................. 135
Table 4.3 Test data specifications..................................................................................... 157
Table 4.4 Error of the estimated EOP using single photo resection. ......................... 169
Table 4.5 One standard deviation of the estimated EOP using single photo resection. . 169
Table 4.6 Image coordinates residuals and errors of the estimated EOP using single photo resection. ........................................................................................................ 170
Table 4.7 Error of the estimated EOP using bundle adjustment (without outlier removal).

Table 4.8 One standard deviation of the estimated EOP using bundle adjustment (without outlier removal).

Table 4.9 Bundle adjustment (Baarda’s data snooping). $F_\alpha$ is F-distribution at $\alpha$ significance level, DOF is degree of freedom.

Table 4.10 Flagged outlier image points after the first bundle adjustment.

Table 4.11 Error of the estimated EOP using bundle adjustment (with outlier removal).

Table 4.12 One standard deviation of the estimated EOP using bundle adjustment (with outlier removal).

Table 4.13 Ground restitution accuracy.
LIST OF FIGURES

Figure 1.1 HRSI stereo image acquisition ................................................................. 5
Figure 1.2 Well-known epipolar geometry of the frame cameras; the left and right image points are constrained on the straight epipolar line. .................................................. 6
Figure 2.1 Coordinate frames used in the collinearity equation. ............................. 19
Figure 2.2 Epipolar geometry for frame cameras showing coplanarity condition. ....... 23
Figure 2.3 Epipolar geometry for frame cameras based on iterative projection......... 24
Figure 2.4 Unit vectors showing the orientation of the left image, right image and the epipolar resampling plane. .................................................................................. 31
Figure 2.5 Epipolar resampling plane...................................................................... 31
Figure 2.6 Frame camera (left) vs. Pushbroom camera (right). ............................ 39
Figure 2.7 Comparison of important epipolar properties in case of (a) frame camera and (b) pushbroom camera. ................................................................. 45
Figure 2.8 Parallel projection model................................................................. 49
Figure 2.9 Epipolar curve generation in the right image from a left image point for pushbroom sensor. ................................................................. 56
Figure 3.1 Epipolar curve shape from RPC; (a) hyperbolic-shaped curve, (b) magnified curve, and (c) shape changed after RPC modification........................................ 59
Figure 3.2 Test result on the epipolar curve pair from RPC; (a) the left image shows the image center and the epipolar curves corresponding to the two selected points in the right image, (b) the right image shows the epipolar curve from the image center of the left image and two selected points from the ground height in the range 0 m and 1,000 m. ...

Figure 3.3 Discrepancy between two epipolar curves for the entire scene in the left image.

Figure 3.4 Discrepancy between two epipolar curves for the local area around the point $p$ in the left image.

Figure 3.5 Discrepancies in column direction around piecewise-generated epipolar curve points.

Figure 3.6 Piecewise epipolar curve generation and image resampling based on RPC.

Figure 3.7 Piecewise epipolar curve pair generation from RPC (for single image point $p$).

Figure 3.8 Piecewise epipolar curve point generation over the entire images for epipolar resampling.

Figure 3.9 Proposed epipolar resampling method from the generated epipolar curve image points.

Figure 3.10 Example of interpolation for the epipolar resampling; (a) original image, and (b) epipolar resampled image.

Figure 3.11 Ground coordinates in a cubic grid and the corresponding image coordinates (Dial and Grodecki, 2002).
Figure 3.12 Distribution of GCP and tie points over the test IKONOS image of Daejeon, Korea.................................................................................................................................................. 77

Figure 3.13 Piecewise generated epipolar curve points in the left image. Upward triangles indicate the starting point locations for the piecewise approach and dots are the piecewise generated epipolar curve points. ........................................................................................................................................... 81

Figure 3.14 Piecewise generated epipolar curve points in the right image. Downward triangles indicate the point locations obtained by RPC-projection from the upward triangles of the left image using the zero ground height. Dots are the piecewise generated epipolar curve points........................................................................................................................................... 82

Figure 3.15 Straight line fitting residuals of each piecewise generated epipolar curve. .. 84

Figure 3.16 Straight line fitting slope of each epipolar curve. All the epipolar curves from the index 1 to 14 show a small slope difference with respect to one another. ............... 85

Figure 3.17 The arranged epipolar curve points of the left image to satisfy the epipolar image condition (the dots and downward pointing triangles are rearranged points from the generated epipolar curve points in the previous section). ........................................................................................................................................... 86

Figure 3.18 The arranged epipolar curve points of the right image that satisfy the epipolar image condition (the dots and downward pointing triangles are rearranged points from the generated epipolar curve points in the previous section). ........................................................................................................................................... 87

Figure 3.19 The epipolar resampled images (a) left (b) right, generated by the proposed algorithm. ........................................................................................................................................................................................................... 91

Figure 3.20 Anaglyph image created by overlaying the epipolar resampled images. ...... 92
Figure 3.21 Subimages at the corners of the anaglyph image showing that the proposed method achieved small y-parallax over the image.......................... 93

Figure 4.1 Image-to-image matching-based georeferencing methods using reference data: (a) a standard ortho-rectified image and DTM; (b) stereo HRSI................................. 98

Figure 4.2 Flowchart of the proposed method of automatic georeferencing of aerial images using stereo satellite images................................................................. 102

Figure 4.3 The Gaussian-blurred (scaled) images at the first octave shown in the left-bottom, and difference of Gaussian (DoG) images are in the right. Using down-sampling by factor of two, the next octave is obtained (Lowe, 2004)................................. 105

Figure 4.4 Maximum and minimum of the DoG images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles) (Lowe, 2004)................................................................. 106

Figure 4.5 Generation of keypoint descriptors (Lowe, 2004)............................................ 109

Figure 4.6 Test images, CASI-1500 (Courtesy of ITRES Research) (a) full scene (b) subset scenes (c) simulated images using scale difference of 0.8, shear of 0.2, rotation by 45 deg and noise at 1%. .................................................................................................. 111

Figure 4.7 SIFT invariance test as a function of shear distortions................................. 113

Figure 4.8 SIFT accuracy test a function of shear distortions........................................ 114

Figure 4.9 SIFT invariance test a function of scale differences...................................... 115

Figure 4.10 SIFT accuracy test a function of scale differences..................................... 116

Figure 4.11 SIFT invariance test a function of image rotations................................... 117
Figure 4.12 SIFT accuracy test a function of image rotations.......................... 118
Figure 4.13 SIFT invariance test a function of salt & pepper noises................... 119
Figure 4.14 SIFT accuracy test a function of salt & pepper noises...................... 120
Figure 4.15 SIFT invariance test a function of intensity differences..................... 121
Figure 4.16 SIFT accuracy test a function of intensity differences..................... 122
Figure 4.17 SIFT invariance test on spectrum difference (band 14 vs. the others).... 124
Figure 4.18 Flowchart of the RANSAC algorithm........................................... 126
Figure 4.19 RANSAC test to attenuate SIFT matching errors from shear distortion.... 130
Figure 4.20 Multi-scale image matching approach between an aerial image and stereo satellite images................................................................. 132
Figure 4.21 Test reference images (a) epipolar resampled IKONOS image (b) subset image.............................................................................................. 157
Figure 4.22 Test aerial images, [A][B] show agricultural fields, [D][E] are forests, [G] contains buildings and [C][F] represent mixed areas................................. 158
Figure 4.23 IKONOS relative spectral response (GeoEye, 2008)............................ 159
Figure 4.24 Coarse matching results based on SIFT + RANSAC with affine equation. 160
Figure 4.25 Fine matching results based on SIFT + RANSAC with collinearity equation. ............................................................................................................. 163
Figure 4.26 Samples of RANSAC-refined matching points..................................... 166
Figure 4.27 Removed outlier points distribution on the IKONOS image in the iterative bundle adjustment with data snooping (triangles: retained ground controls, crosses: 12 removed points).
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFOV</td>
<td>Angular Field Of View</td>
</tr>
<tr>
<td>CCD</td>
<td>Charged Coupled Device</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>DOG</td>
<td>Difference of Gaussian</td>
</tr>
<tr>
<td>DOF</td>
<td>Degree of Freedom</td>
</tr>
<tr>
<td>DTM</td>
<td>Digital Terrain Model</td>
</tr>
<tr>
<td>EOP</td>
<td>Exterior Orientation Parameters</td>
</tr>
<tr>
<td>GCP</td>
<td>Ground Control Point</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GM</td>
<td>Gauss-Markov</td>
</tr>
<tr>
<td>HRSI</td>
<td>High-Resolution Satellite Image</td>
</tr>
<tr>
<td>INS</td>
<td>Inertial Navigation System</td>
</tr>
<tr>
<td>OGC</td>
<td>Open Geospatial Consortium</td>
</tr>
<tr>
<td>RANSAC</td>
<td>RANdom SAmple Consensus</td>
</tr>
<tr>
<td>RFM</td>
<td>Rational Function Model</td>
</tr>
<tr>
<td>RPC</td>
<td>Rational Polynomial Coefficients</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale-Invariant Feature Transform</td>
</tr>
<tr>
<td>UAV</td>
<td>Unmanned Aerial Vehicle</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Two major technological advances redefined the paradigm of geospatial data acquisition and processing for high-accuracy mapping. One involves the launch of many high-resolution commercial satellites, and the other is the advent of large-format aerial digital cameras. Both technologies have shown high performance in terms of data acquisition capability with entirely digital data flow, and excellent imaging quality that provides an unprecedented amount of geospatial information at the same time. However, there are still technical challenges in the data processing algorithms and methods that are required for accurate and robust mapping.

One challenge in the space domain is the determination of epipolar geometry and performing accurate epipolar image resampling, which is essential in achieving a good 3D mapping performance. Note that high-resolution satellite image (HRSI) stereo acquisition with large swath width enables efficient and economic 3D mapping even over inaccessible areas. However, unlike the widely-used aerial frame cameras, the high-resolution satellite systems are based on the pushbroom camera model, whose properties, which are quite different from those of the frame camera, make it difficult to establish the epipolar geometry. Thus, this study proposes a new method of epipolar image resampling
of HRSI, using the Rational Polynomial Coefficients (RPC), a sensor model distributed by most of the commercial operating satellites.

In airborne surveying, an essential step is the establishment of accurate positions and attitudes of the camera at the time of data acquisition, also referred to as sensor georeferencing. In modern systems, it is highly automated by using integrated systems, based on the Global Positioning System (GPS) and the inertial navigation system (INS). Since GPS is a line-of-sight system and is subject to interference and jamming, any loss of GPS lock results in the so-called free inertial navigation of a GPS/INS system, where georeferencing errors increase with time due to drift of the inertial sensors. Moreover, there are remote, inaccessible areas on Earth that lack geodetic and ground GPS infrastructure required for high-accuracy differential GPS data processing often required by mapping applications. Note that there are worldwide differential global positioning services such as Omnistar for the most of the land areas of the Globe, but there are still some areas not covered and the user must have a clear line-of-sight to the local satellite. Consequently, there is a growing need for an alternative solution to GPS/INS-based georeferencing. Based on the premises discussed above, the use of stereo HRSI as a ground control source for automatic aerial image georeferencing is proposed.
1.1 Epipolar resampling of HRSI

Earth observing satellites have been developed and successfully operated for scientific, commercial and military purposes since the 1970’s; early systems offered coarse resolution and long revisit times. A major milestone was reached on March 10th, 1994, when the Presidential Decision Directive 23 was issued by the Clinton Administration, establishing the first formal policy on licensing of commercial remote sensing, which led to the development of the first commercial HRSI, IKONOS-2, offering a 1-m image resolution and large swath width. Moreover, along-track stereo capability allowed, for the first time, for high-resolution 3D topographic mapping on a global scale. Since then, HRSI have been considered a valuable source for creating base maps for various applications, such as weather monitoring, change detection, global water and climate evolution, and other applications in location and navigation purposes, such as GoogleMap or GoogleEarth, car navigation and intelligent transportation systems. Recently launched high-resolution, high-performance satellite imaging systems provide the highest image resolution that is 50 cm, as restricted by the US Government worldwide, these offer revisit times of only a few days, some even have multiple daily acquisitions, and agile sensor orientation capability, so stereo or even multiple image coverage can be obtained. In Table 1.1, the specifications of HRSI are listed in terms of the spatial, temporal, radiometric, and spectral resolution, and swath width. Note that many satellites provide sub-meter resolution with large swath width of more than 10km. In addition, higher performance satellites, such as CARTOSAT-3, EROS-C, and
GeoEye-2, will be launched in a few years. Therefore, the usage of the HRSI is expected to significantly increase.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IKONOS (1999)</td>
<td>0.8</td>
<td>3–5</td>
<td>11</td>
<td>4</td>
<td>11.3</td>
</tr>
<tr>
<td>EROS-A (2000)</td>
<td>1.8</td>
<td>3–4</td>
<td>11</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Quickbird (2001)</td>
<td>0.6</td>
<td>1~3.5</td>
<td>11</td>
<td>4</td>
<td>16.5</td>
</tr>
<tr>
<td>SPOT-5 (2002)</td>
<td>2.5</td>
<td>2~3</td>
<td>8</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>OrbView-3 (2003)</td>
<td>1</td>
<td>~3</td>
<td>11</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>FORMOSAT-2 (2004)</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>CARTOSAT-1 (2005)</td>
<td>2.5</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>ALOS PRISM (2006)</td>
<td>2.5</td>
<td>2~46</td>
<td>8</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>KOMPSAT-2 (2006)</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>EROS-B (2006)</td>
<td>0.7</td>
<td>3~4</td>
<td>10</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>WorldView-1 (2007)</td>
<td>0.5</td>
<td>4.6</td>
<td>11</td>
<td>1</td>
<td>17.6</td>
</tr>
<tr>
<td>CARTOSAT-2 (2008)</td>
<td>0.8</td>
<td>4~5</td>
<td>10</td>
<td>1</td>
<td>9.6</td>
</tr>
<tr>
<td>GeoEye-1 (2008)</td>
<td>0.4</td>
<td>2.8</td>
<td>11</td>
<td>4</td>
<td>15.2</td>
</tr>
<tr>
<td>WorldView-2 (2009)</td>
<td>0.46</td>
<td>3.7</td>
<td>11</td>
<td>8</td>
<td>16.4 ~17.6</td>
</tr>
</tbody>
</table>

Table 1.1 Currently operating high-resolution satellite image (HRSI) systems.

Most high-resolution satellites are designed to provide along track (in-track) stereo acquisition capability since stereo images have been the primary source for 3D topographic mapping and monitoring, i.e. 3D ground coordinates can be computed with the help of sensor models, by identifying conjugate points and objects from stereo images. Note that the satellites’ stereo acquisition in high resolution combined with large swath width provides an unprecedented advantage for 3D mapping and monitoring. As shown
in Figure 1.1, in contrast to the across track stereo images acquired at different times (Figure 1.1(b)), the along track stereo images (Figure 1.1(a)) are acquired at a very small time interval so that there is little spectral difference between the two images. Therefore, it serves as a good source for topographic mapping.

![Image of stereo image acquisition](image)

Figure 1.1 HRSI stereo image acquisition.

Accurate epipolar geometry determination and epipolar image resampling are pivotal to the use of stereo images for accurate 3D topographic mapping. It is well known that the standard procedure for 3D topographic mapping begins with epipolar image resampling, which aligns images to stereo geometry. In a simple stereo image application, using a stereo display on the computer monitor, human operators can easily identify and extract 3D features, such as points of interest, contours, building layers, and roads. This stereo display can be efficiently implemented by epipolar image resampling. More importantly, the epipolar geometry allows for significant reduction of computational load.
for efficient stereo image processing in real-time, by enabling regions for conjugate points to be constrained along a single line.

The pushbroom camera, which is used by most high-resolution satellites, has quite different epipolar geometry than the frame cameras. Note that frame cameras have well-known epipolar geometry, i.e. two conjugate image points lie on the epipolar plane containing the two exposure stations and an object point, so that it produces straight epipolar line on each image paired with the subsequent image, as shown in Figure 1.2.

![Figure 1.2 Well-known epipolar geometry of the frame cameras; the left and right image points are constrained on the straight epipolar line.](image)

In contrast, the pushbroom camera does not produce straight epipolar lines, and the epipolar pair does not exist for the entire scene. It is not easy to derive the mathematical
expression for epipolar geometry, because the pushbroom cameras on the satellite platforms are rigorously modeled using the polynomial EOP (Exterior Orientation Parameters), (see, Lee et al., 2000), it is because given a ground point, the image point coordinates are determined iteratively in the pushbroom camera model. In spite of this, some researchers have already addressed the difficulties associated with the epipolar geometry. Gupta and Hartley (1997) made an assumption of constant velocity and attitude of the satellite platform, and determined that there are no epipoles, and that the epipolar lines are not straight lines from the analysis of the essential matrix. Kim (2000) derived the epipolar curve equation based on Orun and Natarajan’s (1994) sensor model that models the position and yaw angle of the camera using second-order polynomials, while the pitch and roll angles are constant. Kim (2000) also showed the existence of hyperbolic shaped epipolar curve as well as the nonexistence of epipolar curve pairs for the entire scene. These properties make it difficult to establish the epipolar geometry of the pushbroom camera for epipolar image resampling.

Some researchers have adopted approximate models to avoid these problems and to perform epipolar image resampling. Ono et al. (1999) presented an epipolar resampling method for HRSI of high altitude and narrow AFOV (Angular Field Of View) with the approximation of constant velocity and constant attitude of the satellite platform, which led to the 2D affine model proposed by Okamoto et al. (1999). Similarly, Morgan (2004) derived epipolar geometry and resampling equations based on the parallel projection model that is consistent with the 2D affine model in terms of the assumptions and model
equation form. This study showed that near straight epipolar lines can be obtained. However, the assumption of parallel projection requires the transformation from the perspective to parallel projection using prior information of the focal length and roll angle that can be obtained directly from the camera rotation angle or estimated indirectly from GCP (Ground Control Points). On the other hand, the assumption of the constant attitude could not be satisfied for a longer image strips.

The investigation into epipolar geometry and the development of epipolar resampling based on the Rational Function Model (RFM) has been long required, as it is the most widely used sensor model for HRSI since IKONOS-2 has been launched. Most HRSI vendors provide RPC, which are practically the coefficients of the RFM, instead of the rigorous model, following the Open Geospatial Consortium, Inc (OGC) standards (1999). The advantage of using RPC is that users are not required to know the specific information about the satellite sensors (such as, focal length). Furthermore, there is little difference between RPC and the rigorous model for a given height range (Grodbecki, 2001). For example, sub-pixel accuracy can be achieved using RPC when refined with GCP (Dial and Grodecki, 2002; Fraser et al., 2005). High-resolution stereo images, from recently launched satellites, such as GeoEye-1, provide RPC with high positional accuracy of up to 2 meters of circular error at a 90% confidence level (CE90) without the use of GCP, and sub-meter accuracy can be achieved using a bias-compensation RFM model with a single GCP (Fraser and Ravanbakhsh, 2009). In the future, the number of high-resolution satellites will increase, and their positional accuracy is expected to
improve. Consequently using RPC for 3D topographic mapping is promising, and the development of accurate epipolar resampling algorithms from RPC is necessary.

This study proposes a new method for epipolar curve determination and epipolar resampling of HRSI directly from RPC, and not by converting RPC to a simpler sensor models to avoid complex geometry of the pushbroom camera. The proposed method is based on the finding that the conjugate epipolar curve pairs approximately exist for a local area and the global epipolar pairs can also exist if the local pairs are sequentially linked. Then, epipolar image resampling is established by arranging the generated conjugate epipolar curve pair points to satisfy the epipolar resampling image condition. Experimental results for an IKONOS stereo pair showed superior results, achieving much better accuracy than the conventional method.

1.2 Satellite Stereo Imagery-based Georeferencing of Aerial Images

For decades, large-format film cameras were the pillars of airborne surveying, and were considered the most standardized systems in aerial photogrammetry. The early 1990s signaled the transition from analytical to digital photogrammetry, and by the late 1990s, most processes became digital, including feature extraction, terrain model generation and map compilation; note that digital images were obtained by scanning conventional film. It was at the end of the last millennium when multi CCD-based large-format airborne digital camera systems were introduced, and since then the use of these
cameras has been rapidly increasing. Table 1.2 lists widely used large and medium format digital cameras (Jacobsen, 2010).

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Focal length [mm]</th>
<th>CCD size [micron]</th>
<th>Output image [pixels]</th>
</tr>
</thead>
<tbody>
<tr>
<td>UltraCAMXp</td>
<td>Frame</td>
<td>100 for pan 33 for multi</td>
<td>6</td>
<td>17,310 x 11,310 (pan) 5,770 x 3,770 (multi)</td>
</tr>
<tr>
<td>UltraCAMLp</td>
<td>Frame</td>
<td>70 for pan 33 for multi</td>
<td>6</td>
<td>11,704 x 7,920 (pan) 5,320 x 3,600 (multi)</td>
</tr>
<tr>
<td>ADS80</td>
<td>Pushbroom</td>
<td>63</td>
<td>6.5</td>
<td>12,000 (single line)</td>
</tr>
<tr>
<td>DMC</td>
<td>Frame</td>
<td>120 for pan 25 for multi</td>
<td>12</td>
<td>7,680 x 13,824 (pan) 2,048 x 3,072 (multi)</td>
</tr>
<tr>
<td>DMC II 140</td>
<td>Frame</td>
<td>92 for pan 45 for multi</td>
<td>7.2</td>
<td>12,240 x 11,418 (pan) 6,846 x 6,096 (multi)</td>
</tr>
<tr>
<td>DMC II 230</td>
<td>Frame</td>
<td>92 for pan 45 for multi</td>
<td>5.6</td>
<td>15,552 x 14,144 (pan) 6,846 x 6,096 (multi)</td>
</tr>
<tr>
<td>DMC II 250</td>
<td>Frame</td>
<td>112 for pan 45 for multi</td>
<td>5.6</td>
<td>17,216 x 14,656 (pan) 6,846 x 6,096 (multi)</td>
</tr>
<tr>
<td>RMK D</td>
<td>Frame</td>
<td>45</td>
<td>7.2</td>
<td>5,760 x 6,400</td>
</tr>
<tr>
<td>RCD100, 105</td>
<td>Frame</td>
<td>35, 60, 100</td>
<td>6.8</td>
<td>7,216 x 5,412</td>
</tr>
<tr>
<td>DSS322</td>
<td>Frame</td>
<td>40, 60</td>
<td>9</td>
<td>5,436 x 4,092</td>
</tr>
<tr>
<td>AIC</td>
<td>Frame</td>
<td>50</td>
<td>9</td>
<td>5,440 x 4,080</td>
</tr>
<tr>
<td>JAS 150S</td>
<td>Pushbroom</td>
<td>150</td>
<td>6.5</td>
<td>12,000</td>
</tr>
<tr>
<td>Trimble Aerial Camera x2 60</td>
<td>Frame</td>
<td>35/47/60/72</td>
<td>6</td>
<td>8,924 x 6,732</td>
</tr>
<tr>
<td>Trimble Aerial Camera x4</td>
<td>Frame</td>
<td>60/72/100</td>
<td>6</td>
<td>Approximate 17,000 x 12,400</td>
</tr>
<tr>
<td>DIMAC WiDE</td>
<td>Frame</td>
<td>70/120/210</td>
<td>6</td>
<td>13,000 x 8,900</td>
</tr>
<tr>
<td>DIMAC LiGHT</td>
<td>Frame</td>
<td>47/70/120/210</td>
<td>6</td>
<td>8,984 x 6,732</td>
</tr>
</tbody>
</table>

Table 1.2 Aerial digital cameras.
The most notable advantage of these systems, besides the fully digital implementation, is the superior radiometric properties with capability to simultaneously acquire panchromatic, color and near-infrared imagery that allows for significantly better processing performance, as compared to scanned imagery. A critical step in aerial image processing is georeferencing establishing sensor orientation which is usually performed manually using ground control points. Traditionally, it used to be costly and labor intensive, but now it is highly automated by using integrated systems based on GPS/INS (Grejner-Brzezinska, 1999; Mostafa and Schwarz, 2000; Mostafa et al., 2001). The large-format airborne digital camera systems with direct sensor orientation systems enable a large amount of data acquisition and frequent update of geospatial image information with a capability for fast and automated digital processing.

In direct georeferencing by GPS/INS integration, GPS measurements are used to correct INS measurements through an integration algorithm. However, since GPS is a line-of-sight system and is subject to interference and jamming (Carroll, 2001), any loss of GPS lock results in so-called free inertial navigation of a GPS/INS system, where georeferencing errors increase with time due to inertial sensor drifts. Navigation-grade INS can maintain horizontal position accuracy within 100 m through GPS outages of more than 10 minutes. With the lower cost INS, which is common in guided weapons, unmanned air vehicles and general aviation (private) aircraft can only maintain this accuracy for 2 to 3 minutes (Runnalls et al., 2005).
Image-to-image matching-based indirect georeferencing, also referred to as terrain-based navigation is an emerging field which is rapidly gaining research and practical interest and which offers an alternative approach to the georeferencing problem. By means of image matching, existing reference images can provide newly acquired images with ground control information to perform georeferencing. Therefore, reference data quality is one of the key components for the success of georeferencing. Ideal reference data would be images acquired with the same sensor at a similar geometry, at similar time and season as that of the new image; obviously this is rarely the case. Ideally, the sensor used for the reference image should be of high spatial resolution, high temporal resolution, and high positional accuracy.

The image-to-image approach for automatic image co-registration has been studied extensively, see (Wen et al., 2008; Yi et al., 2008; Wong and Clausi, 2007; Le Moigne et al., 2006; Ali and Clausi, 2002; Zhang et al., 2000). Most of these authors focused on two-dimensional co-registration between low-resolution satellite images and airborne imagery using a simple geometric model such as affine geometry and polynomials. Other research utilized a standard ortho-rectified image\(^1\) and an associated low resolution digital elevation model (DEM\(^2\)) to obtain 3D ground control points to georeference satellite images (Rottensteiner et al., 2009; Gianinetto and Scaioni, 2008; Müller et al., 2007). These image-to-image matching approaches using a satellite image were also

---

\(^1\) In the standard ortho-rectified image, only terrain relief is corrected.

\(^2\) A DEM contains elevation information for all surfaces such as buildings and trees.
investigated in the field of navigation to support unmanned aerial vehicle (UAV) navigation (Sim et al., 2002; Conte and Doherty, 2008).

In contrast to the low-resolution image georeferencing, aerial image georeferencing requires accurate 3D ground coordinates. Aforementioned the combination of a standard ortho-rectified image and an associate low resolution DEM has limitations to satisfy this requirement. First, a standard ortho-rectified reference image will typically contaminate the quality of the extracted ground control point due to the still existing relief displacement of the ground features in the image, such as buildings and trees. Second, if a coarse DEM is used, it will lead to inaccurate height information for lack of sufficient spatial resolution. The combination of a ‘true’ ortho-rectified image and a high-resolution DEM could provide accurate 3D ground control information. However, there are practical limitations to the generation and update of a ‘true’ ortho-rectified image for a large area, including the high production cost of a DEM with accurate break-lines. This study proposes the use of stereo HRSI as a ground control source for automatic aerial image georeferencing. The idea was motivated by that stereo HRSI is not subject to the aforementioned effect of relief displacement or the requirement for accurate external

---

3 the apparent leaning outward of objects from the principal point in an aerial image, due to the conical field of view of the camera lens.

4 In the true ortho-rectified image, not only terrain relief, but also the ground features relief, such as buildings, is corrected.
height information, in addition to the attractive properties of HRSI, such as high-positional accuracy for a large swath width and global coverage.

The method proposed in this doctoral research is based on a multi-scale image matching approach utilizing a combination of the SIFT (Scale-Invariant Feature Transform) (Lowe, 1999) and RANSAC (RANdom SAmple Consensus) (Fischler and Bolles, 1981). Aerial images are matched to stereo HRSI to obtain ground coordinates for aerial image matching points. RANSAC, implemented with the collinearity constraint is used to eliminate matching outliers in each aerial image. In the georeferencing step, exterior orientation parameters (EOP) are estimated from the well known bundle adjustment, which is augmented with outlier removal based on Baarda’s data snooping (Baarda, 1968). Experimental results for a strip of aerial digital images showed the potential of this proposed method as an alternative system for automated direct orientation.

1.3 Dissertation outline

This research has two major objectives: (1) epipolar image resampling of stereo HRSI using RPC, and (2) the use of stereo HRSI for automatic georeferencing of aerial digital images.
Chapter 2 describes epipolar geometry and epipolar resampling for commonly used sensor models; namely, the frame camera and pushbroom camera. It also describes the 2D affine model, and RPC. Basic concepts and algorithms pertaining to the latter are introduced. In addition, the challenges in the epipolar geometry and epipolar image resampling in pushbroom camera is addressed.

Chapter 3 discusses the properties of epipolar curves from RPC, followed by the method proposed to determine the epipolar curve for the entire scene. Then, the new method for epipolar image resampling of HRSI is introduced. Experimental results for IKONOS stereo images are presented with discussion.

Chapter 4 explains the novel use of stereo HRSI for automated georeferencing of aerial images. The concept and procedure consisting of robust image matching, outlier removal and georeferencing are presented along with experimental results for an aerial image strip.

Chapter 5 provides a summary, conclusion, and the contributions of this research and suggests some future research directions.
2 OVERVIEW OF EPIPOLAR GEOMETRY AND RESAMPLING

Determination of epipolar geometry and accurate epipolar image resampling are important steps for stereo image processing, including image matching, Digital Elevation Model (DEM) generation and stereo display. Image matching, such as cross-correlation and least squares matching, has been shown to greatly benefit from epipolar geometry to reduce the computation cost as well as mismatches. Even though feature-based matching, such as SIFT (Scale-Invariant Feature Transform, Lowe, 1999), is highly invariant to diverse imaging conditions, i.e., difference in camera position, attitude and intensity, many methods heavily depend on the epipolar constraint to detect and remove outliers in the matching results. For stereoplotting on a photogrammetric workstation, accurate epipolar resampling is pivotal to achieve zero y-parallax and make x-parallax linearly proportional to the ground height. It should be noted that a few pixels of error in each image can be detrimental to ground height determination accuracy.

This chapter presents an overview on the epipolar geometry and the epipolar image resampling of frame camera images, followed by the discussion on the epipolar geometry
of the pushbroom sensor including the rigorous model as well as popular approximate models such as 2D affine and RFM.

2.1 Frame camera

Digital frame cameras use the two-dimensional CCD ( Charged Coupled Device) sensor or sometimes CMOS; either single CCD or multiple CCD sensors for higher resolution and increase in ground coverage. The geometry of frame camera model is based on the well-known collinearity equation, Eq.(2.1), which epipolar line equation can be derived from.

\[
\begin{bmatrix}
  x - x_0 + \Delta x \\
  y - y_0 + \Delta y \\
  -f
\end{bmatrix}
= \lambda M
\begin{bmatrix}
  X - X_L \\
  Y - Y_L \\
  Z - Z_L
\end{bmatrix}
\]  

(2.1)

Where,

- \( x, y \): the coordinates of an image point in the photo coordinate frame (conventionally, \( x \) is along flight direction);
- \( f \): the focal length;
- \( M \): the rotation matrix from the local coordinate frame to the camera coordinate frame which is determined by the attitude of camera at the moment of exposure;
- \( X, Y, Z \): the coordinates of a ground point in the local coordinate frame;
- \( X_L, Y_L, Z_L \): the coordinates of camera position at the moment of exposure in the local coordinate frame;
- \( x_0, y_0 \): the principal point coordinates;
\[ \Delta x, \Delta y \text{ : the camera distortion; } \]
\[ \lambda \text{ : the scale factor (photo scale).} \]

The rotation matrix \( M \) is constructed by using three sequential rotations: roll (\( \omega \)), pitch (\( \phi \)), and yaw (\( \kappa \)) as shown in Eqs. (2.2) and (2.3).

\[
M = M_\kappa M_\phi M_\omega 
\]  \hspace{1cm} (2.2)

\[
M_\omega = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \omega & \sin \omega \\
0 & -\sin \omega & \cos \omega 
\end{bmatrix},
M_\phi = \begin{bmatrix}
\cos \phi & 0 & -\sin \phi \\
0 & 1 & 0 \\
\sin \phi & 0 & \cos \phi 
\end{bmatrix},
M_\kappa = \begin{bmatrix}
\cos \kappa & \sin \kappa & 0 \\
-\sin \kappa & \cos \kappa & 0 \\
0 & 0 & 1 
\end{bmatrix} 
\]  \hspace{1cm} (2.3)

From the left side of Eq.(2.1), the camera distortion parameters and focal length can be separated as a camera matrix \( C \), as shown in Eq.(2.4), where \( C \) stands for calibration.

\[
\begin{bmatrix}
x - x_0 + \Delta x \\
y - y_0 + \Delta y \\
-f
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & -x_0 + \Delta x \\
0 & 1 & -y_0 + \Delta y \\
0 & 0 & -f
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix}
= C \begin{bmatrix}
x \\
y \\
1
\end{bmatrix} 
\]  \hspace{1cm} (2.4)

Where,

\[
C = \begin{bmatrix}
1 & 0 & -x_0 + \Delta x \\
0 & 1 & -y_0 + \Delta y \\
0 & 0 & -f
\end{bmatrix} 
\]  \hspace{1cm} (2.5)
Note that $x$ and $y$ are in the photo coordinate frame, not in the image coordinate frame of pixel unit. Figure 2.1 shows all coordinate frames used in the collinearity equation. The local coordinate frame ($X,Y,Z$) is used for mapping. The camera coordinate frame ($U,V,W$) is centered at the camera exposure position (optical center). $U$ is oriented along the flight direction and $W$ is the actual optic axis, orthogonal to the image plane. $x$ and $y$ are in the photo coordinate frame with an origin at the center of the image and measured in metric units while $r$ and $c$ are in the image coordinate frame, measured from the top-left corner in pixel unit.

\[ \begin{bmatrix} x \\ y \\ f \end{bmatrix} = M \begin{bmatrix} U \\ V \\ W \end{bmatrix} \]

\[ \begin{bmatrix} r \\ c \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} \]

\[ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \]

The coordinates of camera position

\[ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \]

The ground point coordinates

\[ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \]

The ground point coordinates

\[ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \]

The ground point coordinates

Figure 2.1 Coordinate frames used in the collinearity equation.
It is important to understand the transformations between various coordinate frames. The ground point in the local coordinate frame \((X,Y,Z)\) can be transformed into the camera coordinate frame \((U,V,W)\) using the rotation matrix \(M\) of the camera attitude angles as represented by Eq. (2.6).

\[
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix} = M
\begin{bmatrix}
X - X_L \\
Y - Y_L \\
Z - Z_L
\end{bmatrix}
\tag{2.6}
\]

The transformation from the camera coordinate frame \((U,V,W)\) to the photo coordinate frame \((x,y)\) is carried out as Eq.(2.7), which is derived from the collinearity Eq.(2.1) and Eq.(2.6). Note that the equation is simply scaling using the scale factor \(\lambda\).

\[
\begin{bmatrix}
x - x_0 + \Delta x \\
y - y_0 + \Delta y \\
-f
\end{bmatrix} = \lambda
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix}
\tag{2.7}
\]

By dividing the first and the second row with the third row, Eq.(2.8) can be obtained.
If we assume no camera distortion and principal point offset \((x_0 = \Delta x = y_0 = \Delta y = 0)\), Eq. (2.8) becomes Eq. (2.9) which is the transformation from the camera coordinate frame to the photo coordinate frame.

\[
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix} = \frac{-f}{W} S \begin{bmatrix}
U \\
V \\
W
\end{bmatrix}
\]

(2.9)

Where,

\[
S = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1/f
\end{bmatrix}
\]

(2.10)

The conversion between the image coordinate frame \((r, c)\) and the photo coordinate frame \((x, y)\) \(A\) can be expressed generally using a 2D affine transform as shown in Eq. (2.11).

\[
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix} = \begin{bmatrix}
a_1 & a_2 & a_0 \\
b_1 & b_2 & b_0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
r \\
c \\
1
\end{bmatrix} = A \begin{bmatrix}
r \\
c \\
1
\end{bmatrix}
\]

(2.11)
with

\[
A = \begin{bmatrix}
a_1 & a_2 & a_0 \\
b_1 & b_2 & b_0 \\
0 & 0 & 1
\end{bmatrix}
\]  

(2.12)

In case of a CCD in a digital camera, the conversion matrix \( A \) can be typically expressed as Eq. (2.13) using CCD information.

\[
A = \begin{bmatrix}
0 & \rho & -\rho \times (COL / 2 + 0.5) \\
-\rho & 0 & \rho \times (ROW / 2 + 0.5) \\
0 & 0 & 1
\end{bmatrix}
\]  

(2.13)

Where,

\( \rho \): the size of a pixel in CCD in mm;

\( COL, ROW \): the number of pixels in row and column direction of the CCD.

2.1.1 Epipolar line equation

Epipolar geometry is basically explained using the coplanarity condition as depicted in Figure 2.2, that is the condition in which the left and right camera exposure stations \((O_1, O_2)\), an object point \((P)\), and the left and right image points of the object point \((p_1, q_1)\) lie in a common plane. In Figure 2.2, the epipolar plane is defined as a plane
containing the two exposure stations and an object point. Epipolar line is defined as the intersection of the epipolar plane with the left and right photo planes.

Figure 2.2 Epipolar geometry for frame cameras showing coplanarity condition.

Another way of epipolar geometry determination is through the iterative inverse and forward projections, as shown in Figure 2.3. Starting from \( p_l \) on the left image, inverse
projection to a given height and then forward projection to the right exposure station $O_2$, yields an image point on the right image $q_1$. By changing the ground heights, a straight line is formed over the entire right image. By analogy, starting from a right image point on the determined line, a straight line over the whole left image can be formed back. Important property is that any right image point on the line yields the identical epipolar line on the left image. It is known existence of epipolar line pair for entire image.

Figure 2.3 Epipolar geometry for frame cameras based on iterative projection.
In this section, the epipolar line equation is derived from the coplanarity condition.

The vector connecting the ground point and the camera exposure station, Eq.(2.14), can be obtained from the collinearity Eqs.(2.1) and (2.4). See Eq. (2.1) for notation.

\[
\lambda \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix} = M^T C \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.14)
\]

By analogy, expressions for three vectors \( v_1 = \overrightarrow{O_1 O} \), \( v_{i2} = \overrightarrow{O_1 O_2} \), and \( v_2 = \overrightarrow{O_2 O} \) shown in Figure 2.2 are obtained as :

\[
v_1 = M_1^T C_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (2.15)
\]

\[
v_{i2} = \begin{bmatrix} X_{i2} - X_{i1} \\ Y_{i2} - Y_{i1} \\ Z_{i2} - Z_{i1} \end{bmatrix} \quad (2.16)
\]

\[
v_2 = M_2^T C_2 \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (2.17)
\]

Where subscript ‘1’ and ‘2’ indicate the left and the right photo, respectively.
Since the three vectors, \( v_1, v_{12}, \) and \( v_2 \), are on the same plane (epipolar plane), which is the coplanarity condition, the triple product of the vectors should be zero as Eq.(2.18).

\[
\text{Det}(v_1, v_{12}, v_2) = v_1^T \cdot (v_{12} \times v_2) = 0
\]  

(2.18)

Recall that the cross product can be replaced with matrix-vector multiplication as Eq.(2.19).

\[
x \times y = \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}
\]  

(2.19)

Therefore, the cross product part in the coplanarity condition Eq.(2.18) becomes Eq.(2.20).

\[
v_{12} \times v_2 = \begin{bmatrix} 0 & -(Z_{L2} - Z_{L1}) & (Y_{L2} - Y_{L1}) \\ (Z_{L2} - Z_{L1}) & 0 & -(X_{L2} - X_{L1}) \\ -(Y_{L2} - Y_{L1}) & (X_{L2} - X_{L1}) & 0 \end{bmatrix} v_2 = V v_2
\]  

(2.20)

Where,

\[
V = \begin{bmatrix} 0 & -(Z_{L2} - Z_{L1}) & (Y_{L2} - Y_{L1}) \\ (Z_{L2} - Z_{L1}) & 0 & -(X_{L2} - X_{L1}) \\ -(Y_{L2} - Y_{L1}) & (X_{L2} - X_{L1}) & 0 \end{bmatrix}
\]  

(2.21)
The coplanarity condition Eq.(2.18) can be further derived into Eq.(2.22) with the help of Eqs.(2.20), (2.15), and (2.17).

\[
Det(v_1, v_2, v_3) = v_1^T V v_2 = \begin{bmatrix} x_1^T \\ y_1^T \end{bmatrix}^T V \begin{bmatrix} x_2^T \\ y_2^T \end{bmatrix}
\]

\[
= \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} C_1^T \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} F \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = 0
\]

\[
Det(v_1, v_2, v_3) = \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} C_1^T E_2 \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} F \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = 0
\]  \hspace{1cm} (2.22)

With

\[
E = M_1^T V M_2^T
\]  \hspace{1cm} (2.23)

\[
F = C_1^T M_1^T V M_2^T C_2
\]  \hspace{1cm} (2.24)

Where \( M \) is rotation matrix, \( C \) is camera matrix, and \( V \) is in Eq.(2.21).
The matrix $E$ and $F$ are called the *Essential matrix* and the *Fundamental matrix* respectively. Eq.(2.22) can be expressed as Eq.(2.25) if the image coordinate frame is used instead of the photo coordinate frame using Eq.(2.11):

$$[r_1 \ c_1 \ 1]A^tFA_2 \begin{bmatrix} r_2 \\ c_2 \\ 1 \end{bmatrix} = 0 \Rightarrow [r_1 \ c_1 \ 1]F' \begin{bmatrix} r_2 \\ c_2 \\ 1 \end{bmatrix} = 0 \quad (2.25)$$

Where $A$ is a 2D affine transform for the conversion from the image coordinate frame to the photo coordinate frame.

The epipolar line equation can be obtained from the fundamental matrix equation. As it can be seen in Eq.(2.24), the fundamental matrix can be computed directly from the camera information and EOP parameters. Otherwise, if this information is unknown, the elements of the fundamental matrix $f'_{ij}$, can be estimated using Eq.(2.26) from a number of conjugate image points. Note that Eq.(2.26) is obtained from Eq.(2.25).

$$\begin{bmatrix} r_1r_2 & r_2c_2 & r_1 & c_1r_2 & c_1 & r_2 & c_2 & 1 \\ f_{11}' & f_{12}' & f_{13}' & f_{21}' & f_{22}' & f_{23}' & f_{31}' & f_{32}' & f_{33}' \end{bmatrix}^T = 0 \quad (2.26)$$

Where,
\[
F' = A_i^T F A_2 = \begin{bmatrix}
    f'_{11} & f'_{12} & f'_{13} \\
    f'_{21} & f'_{22} & f'_{23} \\
    f'_{31} & f'_{32} & f'_{33}
\end{bmatrix}
\]  

(2.27)

The elements of the fundamental matrix can be obtained by singular value decomposition since it is in the null space of the coefficient matrix, which is the part in front of the dot in Eq.(2.26). With known fundamental matrix, \( F' \), the coefficients of an epipolar line equation on right image can be easily obtained as:

\[
k^T = [r_1 \quad c_1 \quad 1] F' \]  

(2.28)

From Eqs.(2.25) and (2.28), the epipolar line equation can be obtained as Eq.(2.29). As shown, the epipolar equation is in the straight line equation form.

\[
k^T \begin{bmatrix} r_2 \\ c_2 \\ 1 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} r_2 \\ c_2 \\ 1 \end{bmatrix} = 0
\]

\[
\Rightarrow r_2 = \frac{k_2}{k_1} c_2 - \frac{k_3}{k_1} \]  

(2.29)
2.1.2 Epipolar image resampling

Epipolar image resampling removes all y-parallax and leaves only x-parallax, which is proportional to ground elevation. Epipolar lines become parallel to image rows in the epipolar resampled image domain. In Figure 2.2, the resampling can be achieved by aligning the images such that the line connecting \( p_i \) and \( q_i \) is along the image x-direction. Note that two images acquired from an aircraft with constant altitude and attitude will have the same features in the same rows. Therefore, the epipolar resampling can be thought of as projection of each image to the virtual plane which is defined by the constant attitude. The virtual plane is named the *epipolar resampling plane* in the study. How to determine the orientation of the plane is depicted in Figure 2.4. In the figure, ‘^’ denotes the unit vector indicating the orientation, i.e. \( \hat{U}_1, \hat{V}_1, \hat{W}_1 \) shows the orientation of the left image plane and \( \hat{U}_2, \hat{V}_2, \hat{W}_2 \) is for the orientation of the right image plane. The orientation of the epipolar resampling plane \( \hat{U}, \hat{V}, \hat{W} \), can be determined from the orientations of two image planes. Once the orientation is determined, the rotation from each camera coordinate frame to the epipolar camera coordinate frame represent the epipolar resampling transformation as shown in Figure 2.5.
Figures 2.4 Unit vectors showing the orientation of the left image, right image and the epipolar resampling plane.

Figure 2.5 Epipolar resampling plane.
First, the orientation of the epipolar resampling plane $(\hat{U}, \hat{V}, \hat{W})$, can be determined from the EOP of the two cameras. $\hat{U}$ is easily derived as the unit vector connecting the two exposure stations. $\hat{W}$ is oriented to be midway between $\hat{W}_1$ and $\hat{W}_2$ which are the actual optic axes of the two cameras. Note that $\hat{W}$ should be orthogonal to the $\hat{U}$ axis. Finally, the cross product of $\hat{U}$ and $\hat{W}$ becomes $\hat{V}$. See Eqs.(2.30)-(2.32).

\[
\hat{U} = \frac{O_2 - O_1}{\|O_2 - O_1\|} \tag{2.30}
\]

\[
\hat{W} = \frac{\hat{U} \times \left( \frac{\hat{W}_1 + \hat{W}_2}{2} \times \hat{U} \right)}{\left\| \hat{U} \times \left( \frac{\hat{W}_1 + \hat{W}_2}{2} \times \hat{U} \right) \right\|} \tag{2.31}
\]

\[
\hat{V} = \hat{W} \times \hat{U} \tag{2.32}
\]

Where $O_1 = [X_{L1} \ Y_{L1} \ Z_{L1}]$, $O_2 = [X_{L2} \ Y_{L2} \ Z_{L2}]$

The epipolar resampling is the projection of the original image to the epipolar image plane, as shown in Figure 2.4 and Figure 2.5. First, the inverse projection from a left
camera coordinate frame \((U_i, V_i, W_i)\) to local coordinate frame \((X, Y, Z)\) origin at \(O_i\) is expressed using the transposed rotation matrix of the left camera.

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = M_i^T \begin{bmatrix}
U_i \\
V_i \\
W_i
\end{bmatrix}
\]

(2.33)

Where \(M_i\) is the rotation matrix from local coordinate frame to the left camera coordinate frame.

Then, the forward projection from the local coordinate frame \((X, Y, Z)\) to the left epipolar camera coordinate frame \((U, V, W)\) is obtained using the rotation matrix from the direction cosine as shown in Eqs. (2.34)-(2.36).

\[
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix} = \begin{bmatrix}
\cos UX & \cos UY & \cos UZ \\
\cos VX & \cos VY & \cos VZ \\
\cos WX & \cos WY & \cos WZ
\end{bmatrix} \begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}
\]

(2.34)

Where \(\cos UX\) is the cosine of the angle between the \(X\) axis (local coordinate frame) and the \(U\) axis (camera coordinate frame).
\[
\begin{bmatrix}
U
\vline
V
\vline
W
\end{bmatrix} =
\begin{bmatrix}
\hat{U} \cdot \hat{X} & \hat{U} \cdot \hat{Y} & \hat{U} \cdot \hat{Z}
\vline
\hat{V} \cdot \hat{X} & \hat{V} \cdot \hat{Y} & \hat{V} \cdot \hat{Z}
\vline
\hat{W} \cdot \hat{X} & \hat{W} \cdot \hat{Y} & \hat{W} \cdot \hat{Z}
\end{bmatrix}
\begin{bmatrix}
X
\vline
Y
\vline
Z
\end{bmatrix}
\]

Where,

\[
\hat{X} = [1 \ 0 \ 0]^T, \hat{Y} = [0 \ 1 \ 0]^T, \hat{Z} = [0 \ 0 \ 1]^T
\]

Finally, from Eqs.(2.33) and (2.35) the transformation from a coordinates of the left camera frame \((U_1,V_1,W_1)\) to the coordinates of the left epipolar camera frame \((U,V,W)\) becomes Eqs. (2.37).

\[
\begin{bmatrix}
U
\vline
V
\vline
W
\end{bmatrix} =
\begin{bmatrix}
\hat{U} \cdot \hat{X} & \hat{U} \cdot \hat{Y} & \hat{U} \cdot \hat{Z}
\vline
\hat{V} \cdot \hat{X} & \hat{V} \cdot \hat{Y} & \hat{V} \cdot \hat{Z}
\vline
\hat{W} \cdot \hat{X} & \hat{W} \cdot \hat{Y} & \hat{W} \cdot \hat{Z}
\end{bmatrix}
\begin{bmatrix}
U_1
\vline
V_1
\vline
W_1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\hat{U}^T & \hat{V}^T & \hat{W}^T
\end{bmatrix}
\begin{bmatrix}
U_1
\vline
V_1
\vline
W_1
\end{bmatrix}
= M_{EPI}^T
\begin{bmatrix}
U_1
\vline
V_1
\vline
W_1
\end{bmatrix}
\]
\[
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix} = M_{EPI1} \begin{bmatrix}
U_1 \\
V_1 \\
W_1
\end{bmatrix}
\]  

(2.37)

Where,

\[
M_{EPI1} = \begin{bmatrix}
\hat{U}^T \\
\hat{V}^T \\
\hat{W}^T
\end{bmatrix} M_1^T
\]  

(2.38)

By analogy, the transformation from the right camera coordinate frame \((U_2, V_2, W_2)\) to the right epipolar camera coordinate frame \((U, V, W)\) of the right epipolar camera frame can be easily obtained as Eq. (2.39).

\[
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix} = \begin{bmatrix}
\hat{U}^T \\
\hat{V}^T \\
\hat{W}^T
\end{bmatrix} M_2^T \begin{bmatrix}
U_2 \\
V_2 \\
W_2
\end{bmatrix} = M_{EPI2} \begin{bmatrix}
U_2 \\
V_2 \\
W_2
\end{bmatrix},
\]  

(2.39)

Where,

\[
M_{EPI2} = \begin{bmatrix}
\hat{U}^T \\
\hat{V}^T \\
\hat{W}^T
\end{bmatrix} M_2^T
\]  

(2.40)
The transformation from the epipolar camera coordinate frame \((U,V,W)\) to the epipolar image coordinate frame \((r,c)\) can be performed through the epipolar photo coordinate frame \((x,y)\) as Eq.(2.41) from Eqs.(2.11) and (2.9). Note that no camera distortion is assumed in the epipolar camera coordinate frame.

\[
\begin{bmatrix}
  r \\
  c \\
  1
\end{bmatrix} = A_{EPI}^{-1}
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix} = A_{EPI}^{-1} \frac{-f}{W} S_{EPI}
\begin{bmatrix}
  U \\
  V \\
  W
\end{bmatrix}
\]  

(2.41)

with

\[
A_{EPI}^{-1} = \begin{bmatrix}
0 & -1/\rho_{EPI} & ROW_{EPI} / 2 + 0.5 \\
1/\rho_{EPI} & 0 & COL_{EPI} / 2 + 0.5 \\
0 & 0 & 1
\end{bmatrix}
\]  

(2.42)

\[
S_{EPI} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1/f
\end{bmatrix}
\]  

(2.43)

Where,

\[\rho_{EPI}\] : the pixel size of the epipolar resampled image in mm;

\[COL_{EPI}\] : the column size of the epipolar resampled image in pixels;

\[ROW_{EPI}\] : the row size of epipolar resampled image in pixels.

From Eqs. (2.37), (2.7), and (2.11), one can obtain the epipolar camera coordinate frame, \((U,V,W)\) from a left image coordinate \((r_i,c_i)\) as Eq.(2.44).
Therefore, by substituting Eq. (2.44) into Eqs. (2.41), the transformation from the left image coordinates \( (r_1, c_1) \) to the epipolar image coordinates \( (r, c) \) is represented by Eq. (2.45) and (2.46). Note that the scale factor, \( \lambda \), is now included in the denominator \( W \), which is the third element of \( T_{EPI} \left[ r_1 \quad c_1 \quad 1 \right]^T \) in Eq. (2.45).

\[
\begin{bmatrix}
  r \\
  c \\
  1
\end{bmatrix}
= A_{EPI}^{-1} \cdot \frac{-f}{W} S_{EPI} M_{EPI} C_1 A_i \begin{bmatrix}
  r_1 \\
  c_1 \\
  1
\end{bmatrix}
\]  

\[
\Rightarrow \begin{bmatrix}
  r \\
  c \\
  1
\end{bmatrix}
= \frac{-f}{W} T_{EPI} \begin{bmatrix}
  r_1 \\
  c_1 \\
  1
\end{bmatrix}
\]  

(2.45)

Where,

\[
T_{EPI} = A_{EPI}^{-1} S_{EPI} M_{EPI} C_1 A_i
\]

(2.46)
Finally, the epipolar image resampling equation Eq.(2.45) can be expressed as a set of homogeneous coordinates for the image coordinates as shown in Eqs.(2.47) and (2.48). By analogy, the same equation can be obtained for a right image coordinates.

\[
\begin{bmatrix}
R \\
C \\
S
\end{bmatrix} = T_{EPI} \begin{bmatrix}
R_1 \\
C_1 \\
S_1
\end{bmatrix}
\]  

(2.47)

\[
r = \frac{R}{S}, c = \frac{C}{S}, r_1 = \frac{R_1}{S_1}, c_1 = \frac{C_1}{S_1}
\]  

(2.48)

Where,

\((r, c)\): the epipolar resampled image coordinate in pixels;

\((r_1, c_1)\): the left image coordinate in pixels;

\((R, C, S)\): the set of homogeneous coordinates for \((r, c)\);

\((R_1, C_1, S_1)\): the set of homogeneous coordinates for \((r_1, c_1)\).

### 2.2 Pushbroom camera

In contrast to the conventional frame camera which uses a two-dimensional CCD, pushbroom camera adopts an array of one-dimensional CCDs to produce long continuous image strip which is called ‘image carpet’. Figure 2.6 depicts the difference between the frame camera and the pushbroom camera. The frame camera has only one exposure
station for one image, while each image line has its own exposure station in the pushbroom image. Therefore, each image line can be modeled with the collinearity equation for the pushbroom image.

In the section, the epipolar geometry of the pushbroom sensor will be overviewed for three popular models: rigorous model, 2D affine, and RPC.

Figure 2.6 Frame camera (left) vs. Pushbroom camera (right).
2.2.1 Rigorous model

The collinearity equation for the pushbroom camera is identical to that of the frame camera except that $x$ is set to zero because it has one dimensional CCD along $y$ axis, and each image line has its own EOP parameters. Therefore, the collinearity equation for the image line ‘$i$’ is expressed as Eq.(2.49).

$$
\begin{bmatrix}
0 \\
y \\
-f
\end{bmatrix} = \lambda_i M_i \begin{bmatrix}
X - X_{L_i} \\
Y - Y_{L_i} \\
Z - Z_{L_i}
\end{bmatrix}
$$

(2.49)

Where,

- $y$: the photo coordinate ($x$ is along flight direction);
- $f$: the focal length;
- $M_i$: the rotation matrix which is determined by the attitude of the camera at the moment of exposure (from the local coordinate frame to the camera coordinate frame);
- $X,Y,Z$: the ground coordinates;
- $X_{L_i},Y_{L_i},Z_{L_i}$: the coordinate of camera position at the moment of exposure;
- $\lambda_i$: the scale factor.

The number of EOP in the pushbroom camera is six times of the number of image lines and obviously it cannot be estimated individually. Therefore, EOP are modeled based on the platform motion. For example, airborne platforms could be highly affected...
by abrupt changes due to atmospheric turbulence resulting in significant image distortion. For that case, Lee et al. (2000) proposed the Gauss Markov stochastic process to constrain the adjacent image lines. In contrast, polynomial models (spline models) are used for the spaceborne trajectory because the trajectory is smooth and stable compared to airborne trajectories. Two popular EOP models for the spaceborne trajectory are Gupta and Hartley’s (1997) and Orun and Natarajan’s (1994).

Gupta and Hartley’s (1997) rather simple model assumes that the platform is moving in a straight line at constant velocity with a constant attitude as Eq.(2.50).

\[
\begin{align*}
X_{i} &= X_{0} + i\Delta X_{i} \\
Y_{i} &= Y_{0} + i\Delta Y_{i} \\
Z_{i} &= Z_{0} + i\Delta Z_{i} \\
\omega_{i} &= \omega_{0} \\
\phi_{i} &= \phi_{0} \\
\kappa_{i} &= \kappa_{0}
\end{align*}
\]  

(2.50)

Where,

- \( i \): the image line number;
- \( X_{i}, Y_{i}, Z_{i} \): the location of the exposure station of image line \( i \);
- \( \omega_{i}, \phi_{i}, \kappa_{i} \): the rotation angle of image line \( i \);
- \( X_{0}, Y_{0}, Z_{0} \): the location of the exposure station of the first image line;
- \( \omega_{0}, \phi_{0}, \kappa_{0} \): the rotation angle of the first image line;
- \( \Delta X_{i}, \Delta Y_{i}, \Delta Z_{i} \): the linear change of the location of the exposure station.
Gupta and Hartley (1997) showed that epipolar lines for the pushbroom cameras are not straight lines by deriving the fundamental equation as Eq.(2.51) which is a hyperbola equation. Without ephemeris data the fundamental matrix elements can be estimated indirectly from more than 11 conjugate points.

\[
\begin{bmatrix}
  x_1 & x_1 y_1 & y_1 & 1 \\
  x_2 & x_2 y_2 & y_2 & 1
\end{bmatrix}^T F
\begin{bmatrix}
  x_1 & x_1 y_1 & y_1 & 1 \\
  x_2 & x_2 y_2 & y_2 & 1
\end{bmatrix} = 0
\]  

(2.51)

Where,

\[
F = \begin{bmatrix}
  0 & 0 & f_{13} & f_{14} \\
  0 & 0 & f_{23} & f_{24} \\
  f_{31} & f_{32} & f_{33} & f_{34} \\
  f_{41} & f_{42} & f_{43} & f_{44}
\end{bmatrix}
\]

(2.52)

Orun and Natarajan (1994) modeled the position of the sensor and its yaw angle as second-order polynomials while assumed the pitch and roll angle to be constant as expressed in Eq.(2.53). It is considered more accurate, even though it is mathematically complex and computationally expensive.

\[
\begin{align*}
X_{\kappa} &= X_{\kappa_0} + i\Delta X_{\kappa_1} + i^2 \Delta X_{\kappa_2} \\
Y_{\kappa} &= Y_{\kappa_0} + i\Delta Y_{\kappa_1} + i^2 \Delta Y_{\kappa_2} \\
Z_{\kappa} &= Z_{\kappa_0} + i\Delta Z_{\kappa_1} + i^2 \Delta Z_{\kappa_2} \\
\omega_{\kappa} &= \omega_{\kappa_0} \\
\phi_{\kappa} &= \phi_{\kappa_0} \\
\kappa_{\kappa} &= \kappa_{\kappa_0} + i\Delta \kappa_{\kappa_1} + i^2 \Delta \kappa_{\kappa_2}
\end{align*}
\]  

(2.53)
Where,

\( i \): the image line number;

\( X_{L_i}, Y_{L_i}, Z_{L_i} \): the location of the exposure station of image line \( i \);

\( \omega_i, \phi_i, \kappa_i \): the rotation angle of image line \( i \);

\( X_{L_0}, Y_{L_0}, Z_{L_0} \): the location of the exposure station of the first image line;

\( \omega_0, \phi_0, \kappa_0 \): the rotation angle of the first image line;

\( \Delta X_{L_i}, \Delta Y_{L_i}, \Delta Z_{L_i} \): the linear change of the location of the exposure station;

\( \Delta X_{L_2}, \Delta Y_{L_2}, \Delta Z_{L_2} \): the quadratic change of the location of the exposure station;

\( \Delta \kappa_1, \Delta \kappa_2 \): the linear and quadratic change of yaw angle.

Kim (2000) derived the epipolar curve as Eq. (2.54) from the Orun and Natarajan’s model.

\[
y_r = \frac{A_1 x_l + A_2 y_l + A_3}{(A_4 x_l + A_5 y_l + A_6) \sin Q(x_r) + (A_4 x_l + A_5 y_l + A_6) \cos Q(x_r)}
\]  (2.54)

Where,

\( x_r, y_r \): image coordinates in the right image;

\( x_l, y_l \): image coordinate in the left image;

\( A_i \): constants for a given image coordinate \( x_l, y_l \);

\( Q(x_r) \): a quadratic function of \( x_r \).
The equation describes a hyperbola-like curve but in contrast to the epipolar equation of Gupta and Hartley (1997), the parameters cannot be estimated indirectly from the conjugate points. In addition, Kim (2000) showed the nonexistence of epipolar curve pairs for the entire scene in contrast to the frame camera. These properties make it difficult to establish the epipolar geometry of the pushbroom camera for epipolar image resampling. Figure 2.7 shows these two important properties, i.e., the straightness and the existence of the epipolar pair for the frame camera and the pushbroom camera. Figure 2.7(a) is a frame camera, where the left image point $p_i$, generates a straight line on the right image and right image points $q_1, q_2, \cdots$ along the straight line, generate identical straight lines on the left image. The two straight lines on both images are called an epipolar pair. Figure 2.7(b) depicts that pushbroom sensor does not satisfy these two conditions. From the left image point $p_i$, a curve is obtained on the right image (nonstraightness) and right image points $q_1, q_2, \cdots$ along the curve do not produce identical curves back on the left image (non-existence of epipolar pair).
Figure 2.7 Comparison of important epipolar properties in case of (a) frame camera and (b) pushbroom camera.
2.2.2 2D Affine model

2D affine model is based on two main assumptions: the parallel projection and linear trajectory with constant attitude. Since the high-resolution satellites have very narrow AFOV (Angular Field Of View), the projection can be approximated by parallel rather than perspective model. Note that for instance, AFOV of IKONOS is less than 1 degree with an altitude above ground of 680km (Gruen, 2000). In other words, the focal length can be assumed infinite. In addition to the parallel projection, the constant velocity and constant attitude of the sensor platform (Eq.(2.50)) is also assumed for 2D affine model. With these assumptions, the first and the second row of the collinearity equation (Eq.(2.49)) becomes Eq.(2.55).

\[
0 = a_{11}(X - X_{L_i}) + a_{12}(Y - Y_{L_i}) + a_{13}(Z - Z_{L_i})
\]
\[
y = a_{21}(X - X_{L_i}) + a_{22}(Y - Y_{L_i}) + a_{23}(Z - Z_{L_i})
\]

(2.55)

with

\[
\lambda_i M_i = \begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix}
\]

Where,

\[M_i\] the rotation matrix which is determined by the attitude of the camera at the moment of exposure (from the local coordinate frame to the camera coordinate frame);

\[X, Y, Z\] the ground coordinates;
By substituting Eq.(2.50) into the first equation of Eq.(2.55), the image line number \( i \) can be derived into Eq.(2.56).

\[
i = \frac{a_{i1}(X - X_{i_0}) + a_{i2}(Y - Y_{i_0}) + a_{i3}(Z - Z_{i_0})}{a_{i1}\Delta X_{i_1} + a_{i2}\Delta Y_{i_1} + a_{i3}\Delta Z_{i_1}}
\]

(2.56)

Where,

\( i \): the image line number;
\( X_{i_0}, Y_{i_0}, Z_{i_0} \): the location of the exposure station of the first image line;
\( \Delta X_{i_1}, \Delta Y_{i_1}, \Delta Z_{i_1} \): the linear change of the location of the exposure station;

Finally, the 2D affine model, Eq.(2.57), is obtained by replacing the image line number \( i \) by \( x \).

\[
x = A_1X + A_2Y + A_3Z + A_4
\]
\[
y = A_5X + A_6Y + A_7Z + A_8
\]

(2.57)

Where,

\( A_1 \sim A_8 \): the independent 2D affine coefficients;
\( x, y \): the 2D affine image coordinate;
\( X, Y, Z \): the ground point coordinates.
Ono et al. (1999) derived the epipolar line equation using the 2D affine model as Eq.(2.58). The coefficients can be determined from conjugate image points. The 2D affine model satisfies the two epipolar properties: straightness and existence of epipolar pair.

\[ y_r = C_1x_l + C_2x_r + C_3y_l + C_4 \]  

(2.58)

Where \( C_i \) are the epipolar line parameters, \( x_l, y_l \) and \( x_r, y_r \) are the 2D affine image coordinates of the left and right image, respectively.

The 2D affine model is equivalent to the parallel projection model proposed by Morgan (2004). The mathematical expression is shown as Eq.(2.59) and the graphic is shown in Figure 2.8. Note that the notation followed Morgan (2004), therefore, the notation is not consistent with the collinearity equation. In Figure 2.8, bold dotted lines show how the equation is formed. The image coordinates \( [x \ y \ 0]^T \) are expressed as sum of three vectors: \( [\Delta x \ \Delta y \ 0]^T, sR[X \ Y \ Z]^T \) and \( s\lambda R[L \ M \ N]^T \). For more details, refer to Morgan (2004).
Figure 2.8 Parallel projection model.

\[
\begin{bmatrix}
  x \\
  y \\
  0 \\
\end{bmatrix}
= s\lambda R \begin{bmatrix}
  L \\
  M \\
  N \\
\end{bmatrix}
+ s R \begin{bmatrix}
  X \\
  Y \\
  Z \\
\end{bmatrix}
+ \Delta x \begin{bmatrix}
  0 \\
  0 \\
  0 \\
\end{bmatrix}
\]  \hspace{1cm} (2.59)

Where,

\( x, y \): the 2D affine image coordinate;
\( X, Y, Z \): the ground points coordinates;
\( s \): the scale value (i.e. image scale);
\( \lambda \): the distance between the ground point and its image point;
\( L, M, N \): the unit vector of the parallel projection direction;
\( R = R_x R_y R_w \): the rotation matrix from the local coordinate frame to the image coordinate frame; notation \( R \) is used instead of \( M \) here because \( M \) is already used for parallel projection unit vector;
$\Delta x, \Delta y$: the shift values from the image coordinate frame to the local coordinate frame.

After determining the model parameters, i.e., $L, M, N, \Delta x, \Delta y, s$, for the left and right image, the parallel projection parameters of the epipolar images can be determined as Eq.(2.60).

$$\omega_{EPI} = \phi_{EPI} = 0$$

$$\kappa_{EPI} = \arctan\left(\frac{N_1M_2 - M_1N_2}{N_1L_2 - L_1N_2}\right)$$

$$\Delta x_{EPI} = \frac{(\Delta x_1 + \Delta x_2)}{2}, \Delta y_{EPI} = \frac{(\Delta y_1 + \Delta y_2)}{2}, s_{EPI} = \frac{(s_1 + s_2)}{2}$$

$$L_{EPI1} = L_1, M_{EPI1} = M_1, L_{EPI2} = L_2, M_{EPI2} = M_2$$

Where,

$EPI$: denotes epipolar resampled image;

$1, 2$: denote the left and right image, respectively.

Once parallel projection parameters of epipolar image are determined from Eq.(2.60), the affine coefficients for epipolar image resampling from the original images to the epipolar images $B_{ij}$ can be established as Eq.(2.61). In other words, the epipolar resampled image coordinates $(x_{EPI}, y_{EPI})$ can be computed from the original image point coordinates $(x, y)$.
\[ x_{EPI} = B_1 x + B_2 y + B_3 \]
\[ y_{EPI} = B_4 x + B_5 y + B_6 \]  

(2.61)

with

\[
\begin{align*}
B_1 &= S(m_{11} - m_{31} U), \\
B_2 &= S(m_{12} - m_{32} U), \\
B_3 &= \Delta x_{EPI} + S((m_{11} \Delta x + m_{32} \Delta y) U - m_{11} \Delta x - m_{12} \Delta y), \\
B_4 &= S(m_{21} - m_{31} V), \\
B_5 &= S(m_{22} - m_{32} V), \\
B_6 &= \Delta y_{EPI} + S((m_{31} \Delta x + m_{32} \Delta y) V - m_{21} \Delta x - m_{22} \Delta y) 
\end{align*}
\]  

(2.62)

Where,

\[
R_{EPI} = R_{EPI,1} R_{EPI,2} R_{EPI,3}, \quad S = s_{EPI} / s,
\]

\[
R_{EPI} R^T = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}
\]

(2.63)

\[
U = \frac{R_{EPI,11} L + R_{EPI,12} M + R_{EPI,13} N}{R_{EPI,11} L + R_{EPI,12} M + R_{EPI,13} N}
\]

\[
V = \frac{R_{EPI,21} L + R_{EPI,22} M + R_{EPI,23} N}{R_{EPI,21} L + R_{EPI,22} M + R_{EPI,23} N}
\]

The assumptions of 2D affine model and the parallel projection model cannot be satisfied for many cases. First, the assumption of nearly zero AFOV may not be true especially for low altitude satellites with large swath width. In addition, even if the AFOV is small, a transformation from the perspective to the parallel projection has to be applied for better accuracy using Eq.(2.64) before the model parameters are estimated.
Note that the roll angle and the focal length information are required in this step. Second, the assumption of the constant attitude will be restricted for accurate sensor modeling of rather dynamic trajectory which has to be modeled with a higher polynomial order such as Orun and Natarajan’s (1994), due to the important assumption of the 2D affine and the parallel projection is linear trajectory with constant attitude.

\[
y_{\text{parallel}} = \frac{y_{\text{perspective}}}{1 - (\tan \omega) y_{\text{perspective}} / f}
\]  \hspace{1cm} (2.64)

Where,

- \(y_{\text{parallel}}\): the y-coordinate in the image of parallel projection;
- \(y_{\text{perspective}}\): the y-coordinate in the image of regular perspective projection;
- \(\omega\): the roll angle;
- \(f\): the focal length.

### 2.2.3 RPC- Replacement Sensor Model

Since the introduction of IKONOS-2, RPC has been the most widely used sensor model for HRSI; as the model is described in the specifications of OGC (Open Geospatial Consortium, Inc) (1999). Most HRSI vendors provide RPC instead of the rigorous model. Using RFM has an advantage because users are not required to know the specific information on the satellite sensors, while there is little difference between RFM and the rigorous model for a given elevation range (Grodecki, 2001).
The basic RFM equation in the forward projection form, i.e., ground to an image, is expressed as Eq.(2.65). For given coordinates of a ground point \((\varphi, \lambda, h)\), the corresponding image coordinates \((l, s)\) can be computed. The equation is a nonlinear equation of 80 coefficients (RPC). In practice, RFM of 78 coefficients is often used since the first terms in denominators \(b_i\) and \(d_i\) are usually set to one to avoid coefficient scale ambiguity. The inverse projection, i.e., image to the ground, can be solved from this forward RFM when ground height information is given. The algorithm is described in detail in (Grodecki et al., 2004) and also refer to the Appendix A.

\[
Y = \frac{a^T u}{b^T u}, \quad X = \frac{c^T u}{d^T u}
\]

with

\[
U = \frac{\varphi - \varphi_0}{\varphi_s}, \quad V = \frac{\lambda - \lambda_0}{\lambda_s}, \quad W = \frac{h - h_0}{h_s}, \quad Y = \frac{l - L_0}{L_s}, \quad X = \frac{s - S_0}{S_s}
\]

\[
u = \begin{bmatrix} 1 & V & U & W & VU & VW & UW & V^2 \\ U^2 & W^2 & UVW & V^3 & VU^2 & VW^2 & V^2U \\ U^3 & UW^2 & V^2W & U^2W & W^3 \end{bmatrix}^T
\]

\[
a = [a_1, a_2, \ldots]
\]

\[
b = [1, b_2, \ldots]
\]
\[
\begin{align*}
    c &= \begin{bmatrix} c_1 & c_2 & \ldots \end{bmatrix} \\
    d &= \begin{bmatrix} 1 & d_2 & \ldots \end{bmatrix}
\end{align*}
\]

Where,

\begin{align*}
    X, Y & \quad \text{the normalized image coordinates;} \\
    U, V, W & \quad \text{the normalized ground point coordinates;} \\
    \phi, \lambda, h & \quad \text{the geodetic latitude, longitude and ellipsoidal height of ground point;} \\
    l, s & \quad \text{the image line (row) and sample (column) coordinates;} \\
    \phi_0, \lambda_0, h_0, S_0, L_0 & \quad \text{the offset factors for the latitude, longitude, height, sample and line;} \\
    \phi_s, \lambda_s, h_s, S_s, L_s & \quad \text{the scale factors for the latitude, longitude, height, sample and line.}
\end{align*}

Currently there is no explicit form of epipolar curve equation derived from RPC. However, an epipolar curve can be generated in the right image from a left image using the inverse and forward RPC projection by changing the ground height as shown in Figure 2.9, similar to the frame camera in Figure 2.3. A given point \( p \) in the left image is inverse-projected to a ground point of certain height (given) and then forward-projected to the right image. As the ground height changes, the obtained image point set, points between \( q_1 \) and \( q_2 \) in the right image constitute the epipolar curve for an image point \( p \). By analogy left image points between \( p' \) and \( p \) are obtained for the given right image point \( q_1 \) and left image points between \( p \) and \( p'' \) are obtained for the given right image point \( q_2 \). See Appendix A for the forward and inverse RPC projection.
The epipolar properties of RPC are expected to be similar to the rigorous model as depicted in Figure 2.7. There have been few publications on research which attempted to determine the epipolar geometry for the entire scene (Zhao et al., 2008). One of the reasons is that RPC are generally valid only within the ground height range computed from the RPC’s height offset and scale; i.e., their accuracy is not guaranteed outside that region. Therefore, the epipolar curve generation by iterative inverse and forward projection should be valid only within the RPC-established ground height range.

RPC-based accurate epipolar geometry determination and resampling for the entire image are very important because of the increasing popularity of RPC. Therefore, Chapter 3 presents the epipolar curve properties from RPC and a new algorithm is proposed for accurate epipolar geometry determination. Following the study on the epipolar geometry, the proposed epipolar image resampling algorithm is presented with experimental results.
Figure 2.9 Epipolar curve generation in the right image from a left image point for pushbroom sensor.
3  EPIPOLAR IMAGE RESAMPLING OF SATELLITE IMAGES

3.1 Introduction

In this chapter, a new method for conjugate epipolar curve pair determination and epipolar resampling of spaceborne pushbroom images based on RPC is presented (Oh et al., 2010b). First, it is shown that the properties of the epipolar curves from RPC are consistent with those of the rigorous model. Then test results are presented, showing that the local epipolar curve can be approximated by a straight line, and conjugate epipolar curve pairs exist approximately for a local area. Following the test results, the existence of conjugate epipolar curve pairs for the entire scene is investigated when the locally generated epipolar curves are sequentially linked. Based on this investigation, a piecewise epipolar curve determination algorithm and epipolar image resampling are proposed. Image points, consisting of epipolar curve pairs, are generated piecewise in stereo image pairs followed by relocation of the points to satisfy the epipolar image conditions that x-axis of the resampled image should be aligned along the trajectory, y-axis should be orthogonal to the trajectory. Thus, there should be no y-parallax, and the x-parallax is linearly proportional to the ground height.
This chapter is structured as follows. First, the properties of the epipolar curves from RPC are discussed and the proposed method for epipolar curve pair generation for the entire scene is described. Next, the epipolar image resampling method is explained and followed by test results and evaluation.

3.2 Properties of the epipolar curves from RPC

It has been known that the epipolar curve of a pushbroom sensor is a hyperbolic-like shape (Gupta and Hartley, 1997; Kim, 2000). To test what the epipolar curve for the entire image from an RPC looks like using IKONOS imagery, the epipolar curve in the right image from the left image center point is computed from RPC, by incrementing the ground height by 50m from -10,000 meters to +10,000 meters, as depicted in Figure 2.9. As can be seen in Figure 3.1(a) and (b), the generated epipolar curve for the entire image has a hyperbolic-like shape within the image area. Figure 3.1(c) illustrates the case when the epipolar curve changes its shape when the RPC is updated by adjusting RPC error using the ground control information (Dial and Grodecki, 2002; Fraser and Hanley, 2005). Even if the epipolar curves do not seem hyperbolically shaped within the entire image, they are still hyperbolic-like because the hyperbolic shape appears outside the image region in this case.
Figure 3.1 Epipolar curve shape from RPC; (a) hyperbolic-shaped curve, (b) magnified curve, and (c) shape changed after RPC modification.

Next, a straight line fitting to the epipolar curve points in Figure 3.1(c) is performed to demonstrate that epipolar curve become a straight line for a local area. The test is carried out with different ground height ranges. Note that larger ground height range leads to a longer curve on the image as shown in Figure 2.9. Table 3.1 shows the orthogonal distance to the least-squares fitted straight line from each epipolar curve point; i.e., the residuals in direction orthogonal to the straight line. As the height range increases, the deviation increases up to more than five pixels. Note that the epipolar curves for a local area generated using the small height ranges yield very small residuals. Therefore, it can be concluded that the epipolar curve could be approximated with straight lines at sub-pixel accuracy for a local image scene with a given height range of RPC; this is consistent with the result of Kim (2000). Note that RPC are generally valid only within
the ground height range computed from RPC height offset and scale. In other words, their accuracy is not guaranteed outside the region. Given the range of (53m – 385m) in the case of tested RPC, the results in Table 3.1 seem feasible.

<table>
<thead>
<tr>
<th>Height range [m]</th>
<th>0 – 1,000</th>
<th>0 – 2,000</th>
<th>0 – 3,000</th>
<th>0 – 5,000</th>
<th>0 – 10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean [pixel]</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.22</td>
<td>1.76</td>
</tr>
<tr>
<td>Max [pixel]</td>
<td>0.01</td>
<td>0.05</td>
<td>0.15</td>
<td>0.66</td>
<td>5.56</td>
</tr>
</tbody>
</table>

Table 3.1 Straight line fitting orthogonal direction residual of local epipolar curves.

Another important concern about the epipolar curve is that the epipolar curve pairs do not exist for the entire scene. The property was already described in Figure 2.7(b). To investigate this property for RPC, the same procedure is used; i.e., the two image points, $q_1$ and $q_2$ in the right image, which correspond to the left image point $p$, depending on the ground height range, are selected along the epipolar curve, as can be seen in Figure 2.7(b). Then the selected right image points were projected to the left image to generate an epipolar curve back in the left image. Thus, it was investigated if the identical epipolar curve is obtained. Note that the point notations such as $q_1$, $q_2$, $p'$ and $p''$ shown in Figure 2.7(b) are used in the following test.

Figure 3.2 shows the experimental results for the RPC data. The curve at Figure 3.2(b) is the generated epipolar curve on the right image for the center point $p$, of the left image. Two points $q_1$ and $q_2$ along the epipolar curve in Figure 3.2(b) are selected
considering ground heights of 0 m and +1,000 m. The conjugate point of \( p \) should be somewhere between \( q_1 \) and \( q_2 \) along the epipolar curve and should be determined by its true ground height. Then, epipolar curves for \( q_1 \) and \( q_2 \) are generated on the left image using the ground height range (-15,000 m – +15,000 m) to obtain the curve over the entire image and plotted in Figure 3.2(a). At glance, it looks like the identical epipolar curve is obtained in Figure 3.2(a) because of the scale. However, it should be careful that each curve is not a straight line and the two curves are not identical; the difference between the curves is shown and investigated later.

Figure 3.2 Test result on the epipolar curve pair from RPC; (a) the left image shows the image center and the epipolar curves corresponding to the two selected points in the right image, (b) the right image shows the epipolar curve from the image center of the left image and two selected points from the ground height in the range 0 m and 1,000 m.
Figure 3.3 shows the discrepancy between the two curves of Figure 3.2(a) in the column direction. This epipolar curve set is generated in the same way as explained above, except that another set of two points $q_1$ and $q_2$ is selected using the ground height interval 0 m and +2,000 m for comparison. As seen in Figure 3.3, there is not much discrepancy between two epipolar curves around the start point $p$, but the gap between the epipolar curves tends to grow as the distance from the image center increases. The discrepancy in the case of the larger ground height range (0 m 2,000 m) increases faster than in the case of the smaller height range (0 m 1,000 m). Note that the nonexistence of the global epipolar curve pair can also be observed. In Figure 3.4, the discrepancy is plotted only for the area corresponding to the ground height ranges of 0 m–1,000 m and 0 m–2,000 m. The difference is bounded to less than 0.1 pixels which is sufficiently small to allow the two epipolar curves to be approximated as a single epipolar curve. In other words, the two epipolar curves on the left image in Figure 3.2 can be approximated as identical within the given ground height range. The test results show that the epipolar pairs exist locally which is also consistent with the results of a rigorous model (Kim, 2000).
Figure 3.3 Discrepancy between two epipolar curves for the entire scene in the left image.
Figure 3.4 Discrepancy between two epipolar curves for the local area around the point $p$ in the left image.

Next, taking the points $p'$ and $p''$ which are located in the end of the local area where the discrepancy is negligible (see Figure 3.4), the same experiments can be iteratively performed within the entire image area: i.e., following the epipolar curve generation in the right image for the left image point $p'$, two right image points along the curve are selected and used for the epipolar curve generation in the left image. Figure 3.5 shows the computed discrepancy for the entire image area. The dots show the left image
point at every iteration which constitutes the left image epipolar curve, and the plotted curves indicate the computed discrepancy around the point. Even though the discrepancy grows as the process gets closer to the image boundary, the difference is still small enough to ignore. The same results can be obtained if the discrepancy is investigated in the right image. Therefore, it can be concluded that piecewise-generated epipolar curve pairs can be obtained for the entire scene even though their shape is not straight.

Figure 3.5 Discrepancies in column direction around piecewise-generated epipolar curve points.
3.3 Epipolar curve generation and resampling

It was shown in the previous section that the epipolar curve computed from RPC has similar properties to that of the rigorous model and the local epipolar curves that are computed using the ground height range can be approximated with straight lines; thus the epipolar pair exists for local areas. Note that the epipolar curve pairs for entire image can exist only if the curves are generated piecewise. Based on these findings, a piecewise epipolar curve generation and epipolar resampling method is proposed as described in Figure 3.6. Since the piecewise approach provides arrays of image points along the epipolar curves for the entire scene, the obtained points are rearranged to satisfy the epipolar image condition and then the epipolar resampling transformation can be established. Finally, the new RPC for the epipolar resampled images are generated for mapping purposes.
3.3.1 Piecewise epipolar curve generation

Based on the finding that epipolar curve pairs exist for local areas defined by the ground height range, a method based on piecewise epipolar curve generation is proposed, as shown in Figure 3.7.
Figure 3.7 Piecewise epipolar curve pair generation from RPC (for single image point $p$).

Starting from a point in the left image $p$, for example, the center point, the inverse RPC projection to the assumed maximum and minimum heights on the ground is performed to obtain the two ground points which correspond to the intersections of the bold solid line (1\textsuperscript{st} projection) and the two ground planes in Figure 3.7. This ground height range has to be at least equal to or larger than the actual terrain elevation range and small enough to guarantee the existence of the epipolar curve pairs. Then these two ground points are projected to the right image to determine the corresponding image points $q_1$ and $q_2$ by the forward RPC projection (2\textsuperscript{nd} projection). Note that the epipolar line between the two image points $q_1$ and $q_2$ is approximated by a straight line. Now starting from $q_1$ and $q_2$ in the right image, the corresponding two left image points $p'$ and
can be obtained the same way (2nd projection to the ground and 3rd projection to the left image in Figure 3.7). Thus, the epipolar curve between \( p' \) and \( p'' \) and the epipolar curve between \( q_1 \) and \( q_2 \) can be paired. By continuing the projection process, a linear array of image points in the epipolar curve is obtained in the left and right image. They constitute the epipolar curve pair showing the sensor trajectory direction. For any other image point in an image, the epipolar curve can be obtained the same way.

3.3.2 Epipolar resampling

The epipolar resampling consists of three steps: (1) piecewise generation of the epipolar curve points array over the entire image, (2) arrangement of the curve points to satisfy the epipolar image conditions described below, and (3) the transformation for image resampling. The epipolar image has important conditions about the image axis and image parallax; the x-axis of the epipolar resampled image should be aligned along the sensor trajectory while the y-axis should be orthogonal to the trajectory. In terms of parallax, there should be no y-parallax and the x-parallax is linearly proportional to the ground height (Morgan, 2004).

The piecewise epipolar curve generation was introduced in the previous section, where the left image center point was used as the start point. For the purpose of epipolar image resampling, well-distributed epipolar curve points should be generated over the entire image. Therefore, attention should be paid when selecting start points in the left
image to begin the curve generation. Figure 3.8 depicts how to select the start points in the left image and the corresponding image points in the right image. First, the left image center’s epipolar curve which is the trajectory direction is generated by selecting the left image center point as a start point. Then, the orthogonal direction to the center epipolar curve is obtained by straight line fitting (shown in bold dotted arrow). Next, the start points set can be established along the orthogonal direction using some predefined interval (upward triangles), and then epipolar curve pairs for each start point are generated piecewise in the left and right images, as explained in the previous section. The reason for taking this approach is simply that the obtained start point array is orthogonal to the trajectory and thus can be used for establishing the epipolar image’s y-axis. In the case of the right image, the y-axis can be set up with the conjugate points of the left image’s start points (downward triangles), computed using the constant ground height value; for instance, zero height or the minimum ground height. This height choice is only used to determine base x-parallax.
Computed from the start points using the constant ground height (e.g., minimum height)

Start points for epipolar curve points generation

Generated epipolar curve point $\Theta_L$

The center epipolar curve

Left image

Right image

Figure 3.8 Piecewise epipolar curve point generation over the entire images for epipolar resampling.

For epipolar image resampling all piecewise generated points both in the left and right images are relocated to meet the epipolar resampled image conditions. As can be seen in Figure 3.9, images are first aligned by lining up the start points in the left image and the corresponding right image points along the y-axis in the epipolar resampled domain. It should be noted that the shape of the curves has not been straightened yet, meaning that y-parallax still exists. Removal of the y-parallax can be achieved by assigning a constant row coordinate value to each epipolar curve pair in both images. The linear relationship between the x-parallax and the ground height can be attained if the interval between the epipolar curve points is fixed in both images. It is because the epipolar curve points are computed from the fixed height interval in the proposed
piecewise method. The constant interval can be obtained from the mean inter-point distance (between neighboring points along the epipolar curve) in both images.

![Diagram of epipolar resampling method](image)

Figure 3.9 Proposed epipolar resampling method from the generated epipolar curve image points.

The point rearrangement yields many conjugate points between the original imagery and the epipolar resampled imagery for both left and right images. In addition, they are uniformly distributed, as a grid over the entire image. Therefore, by using the conjugate
points the image resampling equation can be formed. In this study, high-order polynomial
transformations Eq.(3.1) are used, because simple models, such as the affine model
cannot account for the curve-to-straight-line transformation.

\[
\begin{align*}
x' &= a_1 + a_2x + a_3y + a_4x^2 + a_5xy + a_6y^2 + \ldots \\
y' &= b_1 + b_2x + b_3y + b_4x^2 + b_5xy + b_6y^2 + \ldots
\end{align*}
\] (3.1)

Where,
- \(x, y\) : the original imagery coordinates of the conjugate points;
- \(x', y'\) : the epipolar resampled imagery coordinates of the conjugate points;
- \(a, b\) : the polynomial transformation coefficients.

As an alternative to polynomial transformation, an interpolation can also be used
without a transformation equation for the epipolar resampling. Since many conjugate
points are obtained in an approximate grid over the imagery, an image point shown in
Figure 3.10(a) can be assigned an epipolar resampled image coordinate through
interpolation from the surrounding grid points, as shown in Figure 3.10(b).
3.3.3 Generation of the RPC for the epipolar resampled imagery

Following the epipolar image resampling, RPC for the epipolar resampled images have to be computed for stereo mapping. The new RPC can be estimated using the virtual GCP by least squares estimation.

Ground coordinates in a cubic grid are generated within the normalized ground coordinate cube range \([-1,+1]\) in each direction of local latitude, longitude and height \((\lambda, \phi, h)\) and forward RPC projected to the original image to obtain the corresponding image points. Now that each epipolar resampled image coordinate for each original image point can be computed by established transformation, RPC for the epipolar resampled images can be estimated from the ground coordinates in the cubic grid and
corresponding epipolar resampled image points. The estimation algorithm proposed by Tao and Yong (2001) can be used as briefly explained in Appendix A.

Figure 3.11 Ground coordinates in a cubic grid and the corresponding image coordinates (Dial and Grodecki, 2002).
3.4 Test and evaluation

3.4.1 Test data specification and preparation

Along track IKONOS stereo images collected on Nov. 11, 2001 are used for testing and evaluation of the proposed method. The data specifications are shown in Table 3.2. The test area is Daejeon, South Korea and the area coverage is from 36.26° to 36.36° North Latitude and from 127.31° to 127.45° East Longitude. The target area height range computed from RPC height offset and scale is (53 m – 385 m). The left image is acquired a minute later than the right image, that is the left image is taken from the South. The product level of each image is level 2 Geo which has a 50 m horizontal positional accuracy at a 90% confidence level. Four GCP and 16 image tie points are acquired and shown in Figure 3.12. GCP are located at the right bottom corner while tie points are well distributed over the entire image.

<table>
<thead>
<tr>
<th>Site</th>
<th>Daejeon, Korea</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Image product level</td>
<td>IKONOS Level 2 Stereo Geo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left Image</td>
<td>Right Image</td>
</tr>
<tr>
<td>Acquisition date</td>
<td>2001-11-19 / 02:19 GMT</td>
<td>2001-11-19 / 02:18 GMT</td>
</tr>
<tr>
<td>Image size</td>
<td>14336×13816</td>
<td>13824× 13816</td>
</tr>
</tbody>
</table>

Table 3.2 Test data specification.
To improve the georeferencing accuracy of the images, the RPC parameters need to be updated using GCP; note that the RPC accuracy is relatively poor at the IKONOS product level 2. Dial and Grodecki (2002) and Fraser and Hanley (2005) described a model to compensate for and adjust the RPC errors as shown in Eq.(3.2). The left sides in the model are shift/drift-compensated image point coordinates using an affine transformation. The right side of Eq.(3.2) is the image point coordinates computed from...
the RPC. The notations were explained in the previous section but repeated for convenience. Since the left side of Eq.(3.2) has six unknowns, at least three GCP are required for full parameter estimation. But for the case of estimating only shift terms $A_0$ and $B_0$, a single GCP is enough. In this study, the IKONOS RPC are updated by estimating only shift terms from four GCP as Toutin (2006) showed that the IKONOS RPC gave similar satisfactory results with the 3D physical model when the RPC is updated by estimating the shift terms. Note that, in contrast, Quickbird imagery needs to be refined with linear functions due to its relief dependency. Table 3.3 presents the RPC residuals before and after the RPC compensation; note that RMS decreased significantly to the sub-pixel level.

\[ \begin{align*}
   l + A_0 + A_1 l + A_2 s &= \frac{a^T u}{b^T u} L_s + L_0 \\
   s + B_0 + B_1 l + B_2 s &= \frac{c^T u}{d^T u} S_s + S_0
\end{align*} \]  

(3.2)

with

\[ u = \begin{bmatrix} 1 & V & U & W & VU & VW & UW & V^2 \\
    U^2 & W^2 & UVW & V^3 & VU^2 & VW^2 & V^2U \\
    U^3 & UW^2 & V^2W & U^2W & W^3 \end{bmatrix}^T \]  

(3.3)

\[ U = \frac{\varphi - \varphi_0}{\varphi_s}, V = \frac{\lambda - \lambda_0}{\lambda_s}, W = \frac{h - h_0}{h_s} \]  

(3.4)

\[ a = [a_1, a_2, \ldots] \]  

(3.5)
\[ b = [1 \ b_2 \ \cdots] \]  \hspace{1cm} (3.6)

\[ c = [c_1 \ c_2 \ \cdots] \]  \hspace{1cm} (3.7)

\[ d = [1 \ d_2 \ \cdots] \]  \hspace{1cm} (3.8)

Where,

\( l, s \): the observed image line (row) and sample (column) coordinates;

\( A_0, A_1, \ldots, B_2 \): the image adjustment parameters (affine transformation). \( A_0 \) and \( B_0 \) are used for compensating shift and the others are for drift;

\( U, V, W \): the normalized ground point coordinates from \( \phi, \lambda, h \);

\( \phi, \lambda, h \): the geodetic latitude, longitude and ellipsoidal height of ground point;

\( \phi_0, A_0, h_0, S_0, L_0 \): the offset factors for the latitude, longitude, height, sample and line;

\( \phi_s, \lambda_s, h_s, S_s, L_s \): the scale factors for the latitude, longitude, height, sample and line.

<table>
<thead>
<tr>
<th>RPC RMS [pixels]</th>
<th>Before the update</th>
<th>After the update</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row</td>
<td>25.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Column</td>
<td>8.3</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Right</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row</td>
<td>1.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Column</td>
<td>5.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 3.3 RPC residuals in the left and right image before and after the RPC update using GCP.
3.4.2 The piecewise epipolar curve generation

Figure 3.13 and Figure 3.14 shows the epipolar curves and the points in the left and right image generated by the proposed method. The height interval (0 m, 1,000 m) is selected such that the interval is large enough to accommodate the ground height range computed from the RPC (53 m – 385 m). First, the center point of the left image is used to generate the center epipolar curve and then the start points are selected along the orthogonal direction to the center epipolar curve 1,000 column pixels apart to ensure image points are well distributed over the entire image. In Figure 3.13, the upward triangles in the image indicate the start point locations and the dots are the piecewise generated epipolar curve points. The downward triangles in Figure 3.14 are the right image start points computed from the left image start points using the minimum ground height.
Figure 3.13 Piecewise generated epipolar curve points in the left image. Upward triangles indicate the starting point locations for the piecewise approach and dots are the piecewise generated epipolar curve points.
Figure 3.14 Piecewise generated epipolar curve points in the right image. Downward triangles indicate the point locations obtained by RPC-projection from the upward triangles of the left image using the zero ground height. Dots are the piecewise generated epipolar curve points.

Figure 3.15 shows the line fitting residual for each epipolar curve obtained above. As shown in Figure 3.13 and Figure 3.14, a number of epipolar curves were obtained for the entire image. The straight line fitting for each epipolar curve was tested to see
straightness of the curve. In Figure 3.15, the line fitting residual of each curve is drawn for the entire image (along the image row). From the residuals up to ± four pixels, it can be seen that the piecewise generated epipolar curve is not a straight line yet, and the higher-order polynomial fitting is expected to achieve sub-pixel accuracy. It can be concluded that the epipolar curve for the entire image, created by the proposed method still cannot be approximated by a straight line. Figure 3.16 shows the slope for each fitted line. In the graph, the epipolar curve index starts from the leftmost line. Note that the angle has a negative value because the row axis has opposite direction to the conventional y-axis (the line can be thought of as being flipped upside down). The angle displays a small change across the epipolar curve.
Figure 3.15 Straight line fitting residuals of each piecewise generated epipolar curve.
3.4.3 Epipolar resampling and accuracy analysis

As explained in 3.3.2, the piecewise generated epipolar curve points are arranged to satisfy the epipolar image conditions; Figure 3.17 and Figure 3.18 shows the arranged points which are the epipolar resampled points for the left and right images, respectively. The dots and downward triangles are the arranged points from the dot and upward...
triangles shown in Figure 3.13 and Figure 3.14. Note that the starting points (triangles) are now aligned in the y-direction. The interval between the dot points in the x-direction is fixed by computing the mean value over the entire image to ensure the linear relationship between the x-parallax and ground height. The y-coordinates of the dot points in the x-direction are set the same for no y-parallax.

![Figure 3.17](image)

Figure 3.17 The arranged epipolar curve points of the left image to satisfy the epipolar image condition (the dots and downward pointing triangles are rearranged points from the generated epipolar curve points in the previous section).
Epipolar resampling is performed using two approaches introduced in the previous section: the polynomial transformation and interpolation. For the polynomial equation, the second and the third order ones are tested. Table 3.4 shows the mean and max residuals when polynomial resampling equations are tested. Residuals were bounded to the sub-pixel level for both polynomial orders but the third-order polynomial shows
slightly better residual values. It can be concluded that the polynomial equations can model the epipolar image resampling quite well.

<table>
<thead>
<tr>
<th>Residuals [pixels]</th>
<th>2nd-order polynomial</th>
<th>3rd-order polynomial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Max</td>
</tr>
<tr>
<td>Left image</td>
<td>0.07</td>
<td>0.44</td>
</tr>
<tr>
<td>Right image</td>
<td>0.08</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 3.4 Residuals of each polynomial epipolar resampling equation.

For y-parallax analysis, 16 tie points were manually measured and 328 tie points were automatically generated based on RPC. In the tie point generation, $20 \times 20 \times 2$ object points in the cubic grid which is shown in Figure 3.11, were projected to the image, and only points within the image area were selected. The computed y-parallaxes for the tie points are listed in Table 3.5. The test result shows that the y-parallax of the measured tie points is approximately at the one pixel level. In contrast, the RPC-generated tie points show almost zero y-parallax which means that nearly zero y-parallax can be obtained if there is no error in the image measurement and RPC is really accurate. Therefore, it can be concluded that the one pixel level y-parallax for the manually measured tie points mainly due to the measurement error and the RPC accuracy.
<table>
<thead>
<tr>
<th>y-parallax [pixels]</th>
<th>16 measured tie points</th>
<th>328 RPC-generated tie points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Max</td>
</tr>
<tr>
<td>Polynomial order</td>
<td>2nd</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>3rd</td>
<td>0.40</td>
</tr>
<tr>
<td>Linear interpolation</td>
<td>0.41</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Table 3.5 y-parallax for tie points by the proposed method.

The performance of the epipolar resampling approach by parallel projection (Morgan, 2004) was compared to the proposed method using the same data set. The parallel projection parameters were indirectly estimated in two different cases. First, four GCP were used for parallel projection parameter estimation; the test results are presented in Table 3.6. The results show a rather large y-parallax up to 19 pixels. The reason is that the four GCP are not evenly distributed over the scene; they are all located in the lower right corner. Therefore, another test was performed where virtually generated GCP over the entire scene from the updated RPC were used for parameter estimation. The y-parallax for 16 measured tie points significantly decreased, but it is still relatively large up to 4.59 pixels. Compared to the parallel projection-based resampling method, the proposed method showing about maximum 1.26 pixels of y-parallax is clearly superior.
Table 3.6 y-parallax for 16 measured tie points estimated by the parallel projection.

Finally, RPC for the epipolar resampled imagery are generated and the ground restitution accuracy for the four GCP is checked. Table 3.7 shows the accuracy of 1 meter and 0.5 meter RMS for horizontal and vertical coordinates, respectively.

Table 3.7 Ground restitution RMS for the proposed method.

The epipolar resampled images by the proposed method are presented side-by-side in Figure 3.19. The two images are overlaid to generate a stereo anaglyph, which can be stereo-viewed using anaglyph glasses as shown in Figure 3.20. Figure 3.21 presents subimages at the corners of the anaglyph image showing small y-parallax over the images.
Small y-parallax can be identified by comparing the row-coordinates of the red color (left-to-right view image) and cyan (right-to-left view image).

Figure 3.19 The epipolar resampled images (a) left (b) right, generated by the proposed algorithm.
Figure 3.20 Anaglyph image created by overlaying the epipolar resampled images.
Figure 3.21 Subimages at the corners of the anaglyph image showing that the proposed method achieved small y-parallax over the image.

3.5 Summary

In this chapter, epipolar geometry of RPC was investigated and the novel approach of using RPC to determine the epipolar curve pair for stereo HRSI pushbroom images was proposed. Then, the accurate epipolar image resampling algorithm was developed and tested for a stereo IKONOS images showing superior results than the conventional method.
Major findings of this study can be listed as:

1. The epipolar curves from RPC have properties similar to the rigorous model; i.e., non-straightness and non-existence of epipolar pair.

2. RPC produces a hyperbolic-shape epipolar curve over the entire image area, but the curve for a local area can be approximated a straight line.

3. Epipolar curve pairs for the entire image do not exist in a stereo HRSI pair, and cannot be obtained using the conventional epipolar curve generation method. However, the epipolar curve pair approximately exists for a local area.

4. The proposed piecewise approach could successfully generate the epipolar curve pairs over the entire image area. Therefore, an epipolar curve over the entire stereo image pair can be generated for any image point using the proposed method.

5. The epipolar curve generated by the piecewise approach is not a straight line.

6. The shape of the epipolar curves within a stereo image pair are not unique; i.e., they have slightly different slopes.

7. The basic idea of the epipolar image resampling algorithm is arranging the epipolar curve pair points to satisfy the important epipolar image conditions, that is, the x-axis of the epipolar resampled image should be aligned along the sensor trajectory while the y-axis should be orthogonal to the trajectory. In addition, there should be no y-parallax and the x-parallax should be linearly proportional to the ground height.

8. The image transformation equation for epipolar resampling requires high order polynomial equations, such as the second or third order.
9. Since the epipolar image resampling is based on the correspondence of numerous epipolar curve points over the stereo images, interpolation techniques can be used for the resampling.

10. Test results showed that near zero y-parallax could be achievable if positional error of RPC is close to zero.
4 AERIAL IMAGES GEOREFERENCING USING STEREO HRSI

4.1 Introduction

This chapter presents a new automatic aerial image georeferencing method based on image-to-image matching to reference satellite data (Oh et al., 2010a). The method utilizes stereo high-resolution satellite images (HRSI) as reference data (ground control) to avoid the effect of relief displacement of the ground features which often appear in standard ortho-rectified images. In addition, the method does not require accurate external height information. Figure 4.1 compares the proposed method to the conventional method of using a combination of standard ortho-rectified imagery and a digital terrain model (DTM), which is often used as a reference data for an image-to-image matching-based georeferencing. Figure 4.1(a) shows the main steps of the standard ortho-rectified image and DTM-based registration. Note that there exists relief displacement of the ground features, such as building and tree in the reference image; the apparent leaning of objects can be observed. In addition, DTM does not contain the height information of the ground features. When the image points (solid triangles on both aerial image and reference image) are obtained by image matching between the aerial and reference images, wrong ground
point information (hollow triangles) is obtained due to the relief displacement, and due to the missing the height information of the ground features in the DTM. In comparison, Figure 4.1(b) depicts the case of using stereo satellite images as reference data. By image matching of the aerial image to the left and right satellite images, the image points (solid triangles and squares) are obtained in each stereo image (reference). By utilizing stereo matching along the epipolar curve, conjugate image points (hollow triangles and squares) can also be obtained in the stereo image pair. The 3D ground restitution of these conjugate points provide correct 3D ground point coordinates, which is not affected by relief displacement of the ground features in the reference data, and does not require external ground height information.
Ortho-rectified satellite images-based geo-referencing (with DTM)

Reference satellite image (ortho-rectified image)

Relief displacement exist (e.g., building roof)

1. Image matching for

Aerial Photo

2. ground control extraction for

3. EOP estimation

(a)

Image matching points

Extracted ground point information

DTM

Ground control error

Figure 4.1 Image-to-image matching-based georeferencing methods using reference data: (a) a standard ortho-rectified image and DTM; (b) stereo HRSI.

continued
Stereo HRSI-based geo-referencing
- the proposed method

1. Image matching for ▲

2. Stereo Matching

3. Intersection

4. EOP estimation

▲ ■ Image points by matching between aerial and reference images (▲ in the left HRSI, ■ in the right HRSI)

▲ □ Image points by HRSI stereo matching

▲ □ Extracted ground point information
Besides the reference data, another requirement of successful georeferencing is the robust image matching, because aerial images and high resolution satellite images tend to have large image differences due to differences in sensors used, acquisition time and season, camera orientation and so on. Therefore, robust and highly invariant feature matching methods are required. For this a multi-scale approach (see Chapter 4.2.4) is proposed based on SIFT (Lowe, 1999) and RANSAC (Fischler and Bolles, 1981) methods. SIFT is a popular point feature extraction and matching method since it was recognized to be very reliable and invariant to changes in image conditions. Note that some modifications to SIFT were developed to make it more effective such as PCA-SIFT (Ke and Sukthankar, 2004), GLOH (Gradient Location-Orientation Histogram) (Mikolajczyk and Schmid, 2005), CSIFT (Abdel-Hakim and Farag, 2006), SR-SIFT (Yi et al., 2008), SURF (Speeded-Up Robust Features) (Bay et al., 2008) and Robust SIFT (Li et al., 2009). However, they are conceptually similar. The major concern in the image matching is the existence of numerous matching outliers which for example can be pruned by RANSAC with an appropriate geometric model. Note that RANSAC is a technique to estimate parameters of a model through iterations from a set of observations that contain outliers.

Figure 4.2 depicts the flowchart of the proposed method of automated aerial image georeferencing using stereo HRSI. First, aerial images are matched to each satellite image in a multi-scale approach to extract aerial image points, e.g., \( n \) points, and the corresponding left and right satellite image points, e.g. \( n_L \) and \( n_R \) points, where
By performing stereo satellite image matching, conjugate satellite image points can be located in the stereo pair. Of course, poor stereo matching points should be pruned at this stage, and it is assumed that \( m \) stereo conjugate remain. Consequently, the number of corresponding aerial image points is also reduced to \( m \). Then, the 3D ground coordinates are computed from the \( m \) stereo conjugate points by ground restitution. Since 3D ground coordinates have been obtained for the corresponding aerial image points, the aerial georeferencing can be carried out indirectly using the collinearity equation. However, there is still high possibility that numerous image matching outliers exist. Therefore, RANSAC is utilized to remove possible image matching outliers by introducing a geometric constraint of collinearity. Following the successful outlier removal, EOP of the aerial images can be finally estimated using the single photo resection method or the bundle adjustment depending on the level of overlap between the aerial images. In the bundle adjustment step, more rigorous outlier removal based on Baarda’s data snooping (Baarda, 1968) is employed as a kind of ‘safety’ because the image overlap can provide useful geometric constraint to filter out the remaining outliers.
Figure 4.2 Flowchart of the proposed method of automatic georeferencing of aerial images using stereo satellite images.
4.2 Image matching between aerial and satellite images

This study proposes a multi-scale approach based on SIFT for efficient yet robust image matching between aerial and satellite images.

Before using SIFT for geospatial image matching, the characteristics, performance, and matching accuracy of SIFT matching need to be investigated. Therefore, in this section, SIFT is briefly introduced first, and its performance is analyzed by testing it for various potential image differences. Then it is shown that the possible matching outliers have to be suppressed by the RANSAC algorithm. Finally, the multi-scale approach based on SIFT is presented.

4.2.1 SIFT matching

Scale Invariant Feature Transform (SIFT) matching is designed to extract invariant features from images and to perform matching. For effective image matching, the features should be invariant to scale, rotation, affine distortion and intensity changes. SIFT feature extraction stages are briefly summarized as following; refer to (Lowe, 2004) for more detail.
Scale-space extrema detection

At this stage, keypoints are extracted using difference of Gaussian filters (DoG) at different image scale. Features invariant to image scale change are searched using different image scales as shown in Figure 4.3. At the first octave, Gaussian blur shown in Eq. (4.1) is used to generate a set of scaled images, then DoG images are obtained. By down-sampling the first octave image by a factor of two, the next octave image is obtained and the same procedure is repeated.

\[ L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (4.1) \]

Where,
- \( \sigma \): the standard deviation of the Gaussian distribution;
- \( L(x, y, \sigma) \): the scale space of an image, blurred with \( \sigma \);
- \( G(x, y, \sigma) \): the Gaussian at the standard deviation \( \sigma \);
- \( I(x, y) \): the input image;
- \(*\): the convolution operator.

The difference of the Gaussian-blurred images, generated using \( \sigma \) and \( k\sigma \), is simply expressed as Eq.(4.2).

\[ D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4.2) \]

Where \( k \) is a constant multiplicative factor for DoG.
Figure 4.3 The Gaussian-blurred (scaled) images at the first octave shown in the left-bottom, and difference of Gaussian (DoG) images are in the right. Using down-sampling by factor of two, the next octave is obtained (Lowe, 2004).

After DoG images across image scales are generated, as shown in the right-bottom of Figure 4.3, keypoints are located by comparing each sample point to its eight neighboring pixels at the same scale, and to nine neighbors in the scales above and below (total 26 neighbors), as shown in Figure 4.4. In the figure, the point marked with ‘X’ is a sample point, and it is compared to the neighbors (marked with circles). If the sample
point is the local minimum or the local maximum of the neighborhood, it is selected as a keypoint.

Figure 4.4 Maximum and minimum of the DoG images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles) (Lowe, 2004).

**Keypoint localization**

The next step is to remove keypoints with low contrast or poorly localized along the edges. At the initial implementation of SIFT, the location and scale of keypoints obtained in the previous step were used as directly obtained at that step. Later, more accurate localization was performed by fitting a 3D quadratic function to the keypoints to determine the interpolated location of the maximum (Lowe, 2004; Brown and Lowe, 2002).

The keypoint with low contrast can be removed using the value of \( D(x, y, \sigma) \). Lowe (2004) used 0.03 as the threshold for \( |D(x, y, \sigma)| \). Removing keypoints poorly localized
along the edges can be done using the 2 x 2 Hessian matrix, obtained by taking second derivatives of \( D \) as shown in Eqs.(4.3). Then, the keypoints not satisfying Eq. (4.4) are removed.

\[
H = \begin{bmatrix}
D_{xx} & D_{xy} \\
D_{xy} & D_{yy}
\end{bmatrix}
\]  \hspace{1cm} (4.3)

with, \( trace(H) = \lambda_1 + \lambda_2 \), \( det(H) = \lambda_1 \lambda_2 \), and \( \lambda_1 = r \lambda_2 \)

Where,

\( \lambda_1, \lambda_2 \): the eigenvalues;

\( r \): the ratio between the eigenvalues.

\[
\frac{trace(H)^2}{det(H)} < \frac{(1 + r)^2}{r}
\]  \hspace{1cm} (4.4)

**Orientation assignment**

For the next step, each keypoint needs to be assigned an orientation. Magnitude and orientation can be computed using pixel difference, as shown in Eqs.(4.5) and (4.6), respectively.

\[
m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}
\]  \hspace{1cm} (4.5)
\[
\theta(x, y) = \tan^{-1}\left( \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right)
\]  \hspace{1cm} (4.6)

Where \( m(x, y) \) is a magnitude, \( \theta(x, y) \) is an orientation, and \( L(x, y) \) is a scale space of an image at the scale of the keypoint.

**Generation of keypoint descriptors**

After orientation is assigned to a keypoint, the keypoint descriptor is computed as a set of orientation histogram in 4 \( \times \) 4 pixel neighborhoods. Gradient magnitude and orientation of each sample point around the keypoint are computed and weighted by a Gaussian window. Then, the information is accumulated into the orientation histogram of 4 \( \times \) 4 subregions. Each histogram has 8 bins and each descriptor has an array of 4 \( \times \) 4 histograms around the keypoint. This leads to a feature vector with 4 \( \times \) 4 \( \times \) 8 = 128 elements. This descriptor is shown in Figure 4.5. Note that the descriptor in Figure 4.5 has an array of 2 \( \times \) 2 histograms around the keypoint. Conventionally the feature vector of 128 elements is normalized and used for SIFT matching.
SIFT matching

SIFT feature matching is performed by comparing each keypoint descriptor from one image to all the keypoint descriptors from the counterpart image. In other words, the matching counterpart is determined when the most similar feature vector is found. Similarity can be obtained by computing a dot product of two feature vectors, e.g. if two features are same, their similarity will be one. One concern is that if a feature vector is the most similar, it does not mean it is a correct matching point. Therefore, SIFT uses a threshold to filter out similar but wrong matching points. SIFT utilizes the ratio of the maximum similarity to the second similarity to obtain only matching points with outstanding similarity.
4.2.2 SIFT performance analysis

SIFT matching performance is analyzed for simulated image differences such as shear distortion, scale, rotation, noise, intensity, and spectral difference. Three test images Figure 4.6(b), which represent building, residential, and flat areas, are subsets from an aerial image shown in Figure 4.6(a), and the simulated image differences are applied to those sub-images shown in Figure 4.6(c). The analysis was carried out for the SIFT matching between the simulated images shown in Figure 4.6(c) and the original images shown in Figure 4.6(a).

The geometric image differences between the original and simulated images are modeled using the affine transform Eq.(4.7), with any combinations of Eqs.(4.8), (4.9) and (4.10).

\[
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix}^T = T \begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}^T
\] (4.7)

Where \( T \) is an affine transformation matrix, \( x,y \) are the image coordinates in the original image, and \( x',y' \) are the image coordinates of the simulated (distorted) image.

\[
Shear : T = \begin{bmatrix}
1 & 0 & 0 \\
\alpha & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\text{ or } T = \begin{bmatrix}
1 & \alpha & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\] (4.8)

Where \( \alpha \) is the shear parameter.
Scale: \( T = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \) 

(4.9)

Where \( s_x, s_y \) are the scale parameter in \( x \) and \( y \) directions, respectively.

Rotation: \( T = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \) 

(4.10)

Where \( \theta \) is the rotation angle.

Figure 4.6 Test images, CASI-1500 (Courtesy of ITRES Research) (a) full scene (b) subset scenes (c) simulated images using scale difference of 0.8, shear of 0.2, rotation by 45 deg and noise at 1%. 

continued
Figure 4.6 continued

SIFT matching analysis: shear

With changing shear parameter $\alpha$, the number of matching points of SIFT matching is counted and plotted in Figure 4.7. The figure shows that as shear distortion $\alpha$ increases, the number of matching points significantly decrease, meaning that SIFT matching does not overcome significant shear distortion. The test images of buildings and residential area show more matching points are obtained than the image of a flat area. Next, the image matching accuracy was investigated since the correct position of the counterpart points is known from the distortion parameter. The matching errors within the five pixels are plotted in Figure 4.8 and the image points with errors larger than five pixels are marked as outliers and counted. The number of outliers out of the total number of matching points is also presented at the caption of each figure. Figure 4.8 demonstrates that with increasing shear distortion, the overall accuracy tends to decrease. Ideally, the
matching error should be zero for the no-distortion case, but the test showed that the errors for some matching points are not zero. The likely reason is due to the scale-space approach of SIFT; i.e., different octave requires down-sample image resampling which might affect accurate localization of keypoints. Similarly, zero error is not obtained when an image is matched to a subset of itself. At shear of 0.0~0.4, no outliers were observed; for shear of 0.6, two outlier matching points were observed.

![Figure 4.7 SIFT invariance test as a function of shear distortions.](image-url)
Figure 4.8 SIFT accuracy test a function of shear distortions.

**SIFT matching analysis: scale**

Figure 4.9 shows the number of matching points as a function of varying scale. Note that aerial images and satellite images usually have different image scales; therefore, image matching should be highly invariant to scale. In the test, smaller scale parameters
mean larger scale difference and the scale ratio of one is for the same scales of both images. Large image scale difference significantly affects the matching result since smaller scale parameters (larger scale difference) tend to yield fewer matching points. In the accuracy test shown in Figure 4.10, the small scale difference tends to insignificantly affect the accuracy while some inaccurate and outlier points are observed in large scale differences such as in the scale ratio of 0.2.

Figure 4.9 SIFT invariance test a function of scale differences.
SIFT matching analysis: rotation

Figure 4.11 presents the number of matching points as a function of varying image rotation. Note that aerial images often have different rotation (orientation) with respect to the reference image due to the flight direction. In Figure 4.11, it is observed that the same
number of matching points is acquired when rotating the image by 90 and 180 deg. The probable reason that 45 and 135 deg rotations provide different results is that 45 and 135 deg rotations require interpolation of pixel values such as cubic convolution, which often result in slightly different pixel values in the output image than the original image. In contrast, 90 and 180 deg rotations do not require the interpolation. In the matching accuracy test shown in Figure 4.12, there were no matching points showing significantly low accuracy. In other words, SIFT matching has good invariance to image rotations.

Figure 4.11 SIFT invariance test a function of image rotations.
SIFT matching analysis: noise

Figure 4.13 presents the SIFT matching results for ‘salt and pepper’ image noise, which is also called impulse noise. This unwanted noise corrupts an image by randomly
appearing as white and black dots superimposed on the image. It is also introduced to the aerial and satellite images due to a noisy channel, errors during the measurement process and during quantization of the data for digital storage (Chahal and Singh, 2010). Therefore, the image matching should be robust to this phenomenon too. In Figure 4.13, it is observed that the number of matching points significantly decreases for the images with significant noise. Figure 4.14 shows that the matching accuracy only slightly degrades, as the noise increases. In other words, the noise does not have much impact on the SIFT matching accuracy.

![SIFT Invariance Test](image)

**Figure 4.13** SIFT invariance test a function of salt & pepper noises.
(a) noise: 2 %: outliers= 0/249  (b) noise 4 %: outliers= 0/143

(c) noise: 6 %: outliers= 0/101  (d) noise 10 %: outliers= 1/60

Figure 4.14 SIFT accuracy test a function of salt & pepper noises.

**SIFT matching analysis: intensity**

The intensity in aerial images and satellite images can be different depending on the image acquisition time, season, and weather. Therefore, invariance to intensity change is also required for robust image matching. Figure 4.15 shows the number of matching
points from the SIFT matching as a function of intensity differences, which was simulated using the power-law equation (Gonzalez and Woods, 2001). Note that gamma is the coefficient in the power-law equation. Gamma of 1 represents no intensity difference and smaller gamma tends to produce the brighter image from an input image. Figure 4.15 shows that the number of matching points decreases as the image intensity difference increases. Matching accuracy slightly degrades as the intensity difference increases in Figure 4.16. The test result indicates that SIFT is highly invariant to the intensity differences.

![Graph showing SIFT invariance test as a function of intensity differences](image)

Figure 4.15 SIFT invariance test as a function of intensity differences.
SIFT matching analysis: spectral difference

Current trend in aerial and satellite sensors is to provide multispectral information, not only the high-resolution panchromatic image; for example, WorldView-2 satellite
(DigitalGlobe) has eight spectral bands. Table 1.2 already listed several aerial cameras with the capability of multispectral data acquisition. Also, hyperspectral images such as CASI-1500 by ITRES are gaining more interest. Therefore, image matching across the spectral bands will be required in many applications. Figure 4.17 shows the result when SIFT matching is performed between the spectral band 14 (878 nanometers wavelength, near-infrared) and bands 1-18 (385~1030 nanometers wavelength) of the test image. The figure clearly shows that the SIFT matching between visible spectrum and near-infrared spectrum produces only a few matching points. Note that the spectral bands from 1 to 10 are in visible spectrums. In addition, the number of matching point decreases significantly as the spectral gap increases. The test results indicate that SIFT cannot handle the large spectral gap. Therefore, it will be very important to select spectral bands of similar spectral ranges for redundant image matching points.
4.2.3 RANSAC

In the SIFT performance analysis, it was shown that SIFT matching may produce inaccurate matching points with the accuracy lower than a few pixel level, which can be significant in the perspective of mapping. In addition, some outlier matching points were observed as error could directly propagate to the quality of aerial image georeferencing. Therefore, these inaccurate matching points should be removed.
RANSAC is a technique to estimate parameters of a model through iterations from a set of observations that contain outliers. Model parameters are estimated from a randomly selected observation subset and then every observation is tested if it fits to the model, and it is added to the consensus set. Through the iteration procedure, a new consensus set is obtained, and a better model is estimated. RANSAC is useful especially when the number of outliers is large, where other robust techniques, such as the least squares residual check or Baarda’s data snooping have practical limitations.

In this section, a brief introduction to RANSAC is presented. For detail, refer to (Fischler and Bolles, 1981; Zuliani, 2009). Figure 4.18 depicts the flowchart of RANSAC. To start the algorithm, the total number of iterations, consensus set size and the error threshold are established. Each iteration randomly samples an observation data set which is called inliner to estimate a model parameter. Now, the model parameter is the temporary consensus set. The remaining observation data were not sampled are referred to as outliers. The errors of outliers with respect to the model parameters based on inliners are computed and an outlier is added to consensus set if the error is less than the established threshold. When all outliers are checked, the size of the consensus set is compared to the selected consensus size to finalize the iteration.
Random sampling of an inliner

\[ \sigma: \text{standard deviation for the measurement error} \]
\[ M: \text{number of iteration} \]
\[ N_{\text{con}}: \text{consensus set size} \]

\[ \text{[Start]} \]

\[ \text{Iteration } i < M \]

\[ \text{Yes} \]

\[ \text{Random sampling of an inliner} \]

\[ \text{Estimate model parameters } p_i, \text{ from the inliner} \]

\[ \text{For each outlier, iterate} \]

\[ i = i + 1 \]

\[ \text{No} \]

\[ \text{[End]} \]

\[ \text{Yes} \]

\[ \text{The consensus set is large enough?} \]

\[ \text{Yes} \]

\[ \text{Compute new model parameters and error to [end]} \]

\[ \text{No} \]

\[ \text{Compute the errors of outliers with respect to the model parameters} \]

\[ j \leq \text{number of outliers} \]

\[ \text{Yes} \]

\[ j = j + 1 \]

\[ \text{No} \]

\[ \text{Add the outlier to the inliner} \]

\[ j = j + 1 \]

\[ \text{No} \]

\[ \text{error} < \text{threshold} \]

\[ \text{Yes} \]

---

Figure 4.18 Flowchart of the RANSAC algorithm.
RANSAC requires appropriate selection of model to fit observation data. Depending on the applications, different model should be selected such as line and plane equation, homography, fundamental matrix and affine model, etc.

Since tests are performed at each iteration to determine if observation data point is an inliner, a threshold $t$ should be established. In the algorithm, a Gaussian error with zero mean and standard deviation $\sigma$ is assumed for the observation error. The sum of squared errors of data to fit a model $d^2$ follows the chi-squared distribution with $m$ degrees of freedom. Therefore, the test statistic for the null hypothesis is:

$$ H_0 : d^2 < t^2 \text{ vs. } H_a : d^2 \geq t^2 $$  \hspace{1cm} (4.11)

$$ T = \frac{d^2}{\sigma^2} \sim F_{m}^{-1}(\alpha) $$  \hspace{1cm} (4.12)

Where,

- $H_0, H_a$ : the null and alternative hypotheses, respectively;
- $t^2$ : the threshold squared to determine if observation data is an inliner;
- $d^2$ : the sum of squared error of observation data fit to a model;
- $\alpha$ : the significance level corresponding to probability that the sample data point is an inliner;
- $F_{m}^{-1}(\alpha)$ : the inverse cumulative chi-squared distribution with degree of freedom $m$.

The threshold squared $t^2$ is determined from a pre-selected established $\sigma$ based on prior knowledge of data quality as shown in Eq.(4.13).
\[ t^2 = F_m^{-1}(\alpha)\sigma^2 \]  

(4.13)

Where \( \sigma \) is the standard deviation of the observation error

In a case when \( \alpha \) is chosen as 0.95 so that there is 95% probability that the data point is an inliner, thresholds and degree of freedom for some applications are listed in Table 4.1.

<table>
<thead>
<tr>
<th>Application</th>
<th>Degree of Freedom, ( m )</th>
<th>( t^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line fitting</td>
<td>2</td>
<td>( 5.99\sigma^2 )</td>
</tr>
<tr>
<td>Plane fitting</td>
<td>3</td>
<td>( 7.81\sigma^2 )</td>
</tr>
<tr>
<td>Affine transformation</td>
<td>6</td>
<td>( 12.59\sigma^2 )</td>
</tr>
<tr>
<td>Single Photo Resection</td>
<td>6</td>
<td>( 12.59\sigma^2 )</td>
</tr>
<tr>
<td>Fundamental matrix</td>
<td>7</td>
<td>( 14.07\sigma^2 )</td>
</tr>
<tr>
<td>Homography</td>
<td>8</td>
<td>( 15.51\sigma^2 )</td>
</tr>
</tbody>
</table>

Table 4.1 Some examples of RANSAC threshold with probability 95%.

Another important task in RANSAC is to establish a threshold on the number of sample selections because it is not feasible to test every possible sample combination. Let us suppose that each random selection draws \( n \) samples from \( N \) observation data, and \( \omega \) is the probability that a sample is an inliner. Then, the probability that all samples belong
to the inliner is $\omega^n$. Conversely, $(1-\omega^n)$ becomes the probability that the samples contain outliers. If total $M$ random selections are made, $(1-\omega^n)^M$ is the probability that any sampling contains outliers. In other words, we never pick a good sample set. This should be small such as 0.1%. Therefore, the number of sample selections, $M$, is determined using Eq. (4.14).

\[(1-\omega^n)^M = 1-p \Rightarrow M = \log(1-p)/\log(1-\omega^n) \quad (4.14)\]

Where,

- $\omega$: the probability that a sample is an inliner;
- $n$: the number of sample points per random selection;
- $M$: the number of selections;
- $p$: the probability that we can avoid the case when we never pick a good sample set.

As a test, affine model-based RANSAC (6 DoF) was applied to attenuate an inaccurate SIFT matching points for the shear distortion as shown in Figure 4.8. In the test, the RANSAC parameters of $\sigma = 0.5$ pixel, $\alpha = 99\%$, $n = 3$, $p = 0.99$ are used. A number of low-accuracy matching points are observed in Figure 4.19(a), but Figure 4.19(b) shows that RANSAC successfully pruned inaccurate matching points.
4.2.4 Multi-scale approach

In addition to robustness and accuracy of image matching, efficiency is also an important criterion for practical applications since geospatial images acquired for mapping purposes or for terrain-referenced navigation are typically large in size. Therefore, a multi-scale approach (also called hierarchical approach, Wong and Clausi, 2007; Zhang et al., 2000) consisting of coarse and fine image matching between aerial and satellite images is proposed, as shown in Figure 4.20.

By utilizing coarse matching, an aerial image is approximately located and coregistered to satellite images and the region of interests (ROI) is obtained in satellite images which the aerial image will be fine-matched. The coarse matching utilizes down-sampled images. Taking the average is the simplest method of low-pass filtering, but in this
study, the Gaussian image pyramid is generated from each aerial and satellite image to attenuate aliasing effects (Fosyth and Ponce, 2000). The top of the image pyramid should have only the low-frequency information. Then, Gaussian down-sampled images of similar spatial resolution are SIFT-matched to each other and outliers in the matching results are removed using RANSAC. Since the down-sampled images are often low-resolution, not significantly affected by relief displacement, the affine model should be satisfactory to model the transformation. Commercial software such as ERDAS Imagine AutoSync requires that at least two conjugate points to be measured by users to orient images and to constrain automatic image matching. In this case, the coarse matching step would not be required because the manually measured data are available. In addition, when external EOP parameters exist, even though they are not accurate enough, the information will be useful for the initial localization or ROI determination. However, in this study, the coarse matching was activated since no external information is assumed.

Following the coarse matching, the fine matching is performed between the aerial image and the selected region of satellite images (each stereo image) to obtain control points from the satellite images. Note that aerial images, which usually have higher spatial resolution, need to be down-sampled close to the resolution of the satellite images before matching, since resolution difference is one of the factors affecting SIFT matching performance, as shown in the performance analysis. For example, if the aerial image has 25cm resolution and satellite images are of 1m resolution, pyramid level 3 of the aerial image is selected. The use of the down-sampled image has another benefit of significantly
reducing computational load in image matching. In contrast to the coarse matching, the outliers in the fine matching results should be suppressed using a rigorous model, such as the collinearity equation exploiting well-calibrated camera information because relief displacement needs to be handled.

Figure 4.20 Multi-scale image matching approach between an aerial image and stereo satellite images.

### 4.3 Stereo matching between stereo satellite images

Since the geometric and spectral difference between the along track stereo satellite images are rather small and the epipolar geometry is well established as studied in Chapter 3, there is no need for SIFT or other sophisticated and computationally expensive feature
matching method. The well known normalized cross-correlation can be used for the stereo image matching using Eq.(4.15).

\[
c = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (A_{ij} - \overline{A})(B_{ij} - \overline{B})}{\sqrt{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} (A_{ij} - \overline{A})^2\right) \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (B_{ij} - \overline{B})^2\right)}}
\]  

(4.15)

Where,

\( A \): the template target image, size of \( r \times c \);  
\( B \): a subarray of template source image of size equal to that of \( A \);  
\( A_{ij}, B_{ij} \): the image intensity of \( A \) and \( B \) at row \( i \) and column \( j \), respectively;  
\( \overline{A}, \overline{B} \): the average of all intensity values in image \( A \) and \( B \), respectively.

Even if the accurate epipolar image resampling significantly reduces the search space to the epipolar line, there is still computational load for any moving window approach. Therefore, the convolution theorem is utilized for fast matching performance. From Eq.(4.15), it can be seen that \( \overline{B} \) can be computed in the frequency domain by FFT, i.e. convolving \( B \) with the window with all elements equal to one. In other words, if \( \overline{B} \) is computed over the entire source image in advance, before the correlation computation, the computational load is significantly reduced.

By decomposing the numerator of the normalized cross-correlation, Eq.(4.15), we can obtain Eq.(4.16). Then, it is identified that \( \sum_{i=1}^{n} \sum_{j=1}^{n} (A_{ij} - \overline{A}) \) can be computed once for each
template target image, and \( \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij} (A_{ij} - \bar{A}) \) can be computed in the frequency domain by convolving \( B \) with \( (A_{ij} - \bar{A}) \).

\[
\sum_{i=1}^{n} \sum_{j=1}^{c_i} \left[ (A_{ij} - \bar{A})(B_{ij} - \bar{B}) \right] = \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij} (A_{ij} - \bar{A}) - \bar{B} \sum_{i=1}^{n} \sum_{j=1}^{c_i} (A_{ij} - \bar{A})
\]  

(4.16)

The decomposition of the denominator of Eq.(4.15) yields Eq.(4.17). Note that

\[
\sum_{i=1}^{n} \sum_{j=1}^{c_i} (A_{ij} - \bar{A})^2 \]

can be computed once for each template target image, and the summation term, \( \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij} \) and \( \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij}^2 \) can be computed by convolving \( B \) or \( B^2 \) with a window that has all elements equal to one.

\[
\left[ \sum_{i=1}^{n} \sum_{j=1}^{c_i} (A_{ij} - \bar{A})^2 \right] \left[ \sum_{i=1}^{n} \sum_{j=1}^{c_i} (B_{ij} - \bar{B})^2 \right] = \left[ \sum_{i=1}^{n} \sum_{j=1}^{c_i} (A_{ij} - \bar{A})^2 \right] \left[ \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij}^2 - \bar{B} \sum_{i=1}^{n} \sum_{j=1}^{c_i} B_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{c_i} \bar{B}^2 \right]
\]  

(4.17)

Performance comparison to a brute-force search was carried out for three test images in Matlab, and the result is presented in Table 4.2. Significant computational cost reduction could be identified.
<table>
<thead>
<tr>
<th>Test image size [pixels]</th>
<th>Normalized cross-correlation run time [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>244×329</td>
<td>4008×2672</td>
</tr>
<tr>
<td></td>
<td>9044.8</td>
</tr>
<tr>
<td></td>
<td>32.96</td>
</tr>
<tr>
<td>244×329</td>
<td>1298×474</td>
</tr>
<tr>
<td></td>
<td>154.91</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
</tr>
<tr>
<td>22×94</td>
<td>713×611</td>
</tr>
<tr>
<td></td>
<td>18.35</td>
</tr>
<tr>
<td></td>
<td>1.77</td>
</tr>
</tbody>
</table>

Table 4.2 FFT based normalized cross-correlation matching performance.
4.4 Georeferencing

4.4.1 Single photo resection

Single photo resection, called space resection, is a method of determining the six EOP of a single photo (Wolf and Dewitt, 2000). At least three control points with known 3D ground coordinates are required to be imaged in the photograph, not aligned on a straight line. Note that two equations are formed for each control point. The collinearity equation is already shown in Chapter 2.1, but the condition equation is repeated in Eqs.(4.18) for convenience.

\[
F_x = x - x_0 + \Delta x + f \frac{U}{W} = 0 \\
F_y = y - y_0 + \Delta y + f \frac{V}{W} = 0
\]

\[
U \\
V \\
W
\]

\[
= M
\left[
\begin{array}{c}
X - X_L \\
Y - Y_L \\
Z - Z_L \\
\end{array}
\right]
\]

Where,

\[
x, y: \text{ the coordinates of an image point in the photo coordinate frame} \\
\text{(conventionally, } x \text{ is along flight direction)};
\]
\( f \): the focal length;
\( M \): the rotation matrix from the local coordinate frame to the camera coordinate frame which is determined by the attitude of camera at the moment of exposure;
\( X, Y, Z \): the coordinates of a ground point in the local coordinate frame;
\( X_L, Y_L, Z_L \): the coordinates of camera position at the moment of exposure in the local coordinate frame;
\( x_0, y_0 \): the principal point coordinates;
\( \Delta x, \Delta y \): the camera distortion;
\( \lambda \): the scale factor indicating photo scale.

The rotation matrix consists of sequential rotations around each axis; roll \((\omega)\), pitch \((\phi)\), and yaw angle \((\kappa)\) as Eq.(4.20).

\[
M = M_\kappa M_\phi M_\omega \tag{4.20}
\]

\[
M_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{bmatrix}, M_\phi = \begin{bmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{bmatrix}, M_\kappa = \begin{bmatrix} \cos \kappa & \sin \kappa & 0 \\ -\sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.21}
\]

With the assumption that the ground control coordinates are fixed, i.e. accurate enough, the collinearity equation Eq.(4.18), is linearized with respect to the EOP parameters \((\omega, \phi, \kappa, X_L, Y_L, Z_L)\) as Eq.(4.22).
\[
\begin{align*}
F_x &= F^0_x + \frac{\partial F_x}{\partial \omega} d\omega + \frac{\partial F_x}{\partial \phi} d\phi + \frac{\partial F_x}{\partial \kappa} d\kappa + \frac{\partial F_x}{\partial X_L} dX_L + \frac{\partial F_x}{\partial Y_L} dY_L + \frac{\partial F_x}{\partial Z_L} dZ_L + e_x = 0 \\
F_y &= F^0_y + \frac{\partial F_y}{\partial \omega} d\omega + \frac{\partial F_y}{\partial \phi} d\phi + \frac{\partial F_y}{\partial \kappa} d\kappa + \frac{\partial F_y}{\partial X_L} dX_L + \frac{\partial F_y}{\partial Y_L} dY_L + \frac{\partial F_y}{\partial Z_L} dZ_L + e_y = 0
\end{align*}
\] (4.22)

The partial derivatives with respect to each EOP parameter are derived as shown in Eq.(4.23) by taking derivatives of Eq.(4.18). (Mikhail et al., 2000)

\[
\begin{align*}
\frac{\partial F_x}{\partial X_L} &= f \frac{\partial U}{\partial X_L} - \frac{U \partial W}{W^2} = f \left( \frac{\partial U}{\partial X_L} - \frac{U \partial W}{W \partial X_L} \right), \\
\frac{\partial F_y}{\partial X_L} &= f \left( \frac{\partial U}{\partial X_L} - \frac{U \partial W}{W \partial X_L} \right), \\
\frac{\partial F_x}{\partial Y_L} &= f \left( \frac{\partial U}{\partial Y_L} - \frac{V \partial W}{W \partial Y_L} \right), \\
\frac{\partial F_y}{\partial Y_L} &= f \left( \frac{\partial V}{\partial Y_L} - \frac{V \partial W}{W \partial Y_L} \right), \\
\frac{\partial F_x}{\partial Z_L} &= f \left( \frac{\partial U}{\partial Z_L} - \frac{V \partial W}{W \partial Z_L} \right), \\
\frac{\partial F_y}{\partial Z_L} &= f \left( \frac{\partial V}{\partial Z_L} - \frac{V \partial W}{W \partial Z_L} \right), \\
\frac{\partial F_x}{\partial \omega} &= f \left( \frac{\partial U}{\partial \omega} - \frac{U \partial W}{W \partial \omega} \right), \\
\frac{\partial F_y}{\partial \omega} &= f \left( \frac{\partial V}{\partial \omega} - \frac{V \partial W}{W \partial \omega} \right), \\
\frac{\partial F_x}{\partial \phi} &= f \left( \frac{\partial U}{\partial \phi} - \frac{U \partial W}{W \partial \phi} \right), \\
\frac{\partial F_y}{\partial \phi} &= f \left( \frac{\partial V}{\partial \phi} - \frac{V \partial W}{W \partial \phi} \right), \\
\frac{\partial F_x}{\partial \kappa} &= f \left( \frac{\partial U}{\partial \kappa} - \frac{U \partial W}{W \partial \kappa} \right), \\
\frac{\partial F_y}{\partial \kappa} &= f \left( \frac{\partial V}{\partial \kappa} - \frac{V \partial W}{W \partial \kappa} \right)
\end{align*}
\] (4.23)

from Eq.(4.19),
\[
\begin{align*}
\frac{\partial}{\partial X_L} \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= M \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \\
\frac{\partial}{\partial \omega} \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= M \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix}, \\
\frac{\partial}{\partial Y_L} \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= M \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix}, \\
\frac{\partial}{\partial \phi} \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= M \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix}, \\
\frac{\partial}{\partial K} \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= M \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix},
\end{align*}
\]

Next, the linearized Eq.(4.22) is arranged according to the Gauss Markov model (Schaffrin, 2007) as Eq.(4.25) which can be solved for the EOP parameter estimates. If \( n \) ground control points are used, then the size of the observation vector and the design matrix will be \( 2n \) and \( 2n \times 6 \), respectively. In the equation, superscript ‘0’ denotes the initial values of the EOP parameters and the subscript indicate the ground point number.

\[
y = A \xi + e, \quad e \sim \left( 0, \sigma_0^2 P^{-1} \right)
\]

Where \( e \) is random error, \( P \) is weight matrix for observation, and often set to be the identity matrix, \( \sigma_0^2 \) is the variance component.
The estimate and their dispersion for the increment of the EOP parameters are computed by the least-squares as shown in Eqs. (4.29) and (4.30), and the residual and the variance component estimates are obtained using Eqs. (4.31) and (4.32), respectively. Note that EOP estimates are obtained by iterations until the solution converges.
\[ \hat{\xi} = \left( A^T PA \right)^{-1} A^T Py \]  
(4.29)

\[ D_{\{\hat{\xi}\}} = \sigma_0^2 \left( A^T PA \right)^{-1} \]  
(4.30)

\[ \sim \]  
(4.31)

\[ \hat{\sigma}_0^2 = \frac{\sim}{2n-6} \]  
(4.32)

Where \( n \) is number of ground control image points.

### 4.4.2 Bundle adjustment

Similarly to a single photo resection, the fundamental equation of bundle adjustment is the collinearity equation. The difference is that in bundle adjustment multiple photos with overlaps are adjusted simultaneously by using tie points in the overlap area. Following the tie points acquisition, ground coordinates of the tie points are computed approximately from prior information or single photo resection method. Then they are regarded as GCP with lower precision (lower adjustment weight) in bundle adjustment. Note that if one tie point pair is obtained in the overlap area of two adjacent photos, the number of image points will be two, and they yield four collinearity equations, i.e., one...
image point yields two collinearity equations. Details on the bundle adjustment can be found in Mikhail et al. (2000) and Wolf and Dewitt (2000).

In the bundle adjustment, in addition to the linearization of the collinearity equation with respect to EOP parameters of each photo, the collinearity equation is often linearized further with respect to GCP coordinates as shown in Eq.(4.33). In the equation, superscript ‘0’ denotes the initial values of the EOP parameters and the GCP coordinates. If $p$ photos and $n$ image points (including tie points) for $m$ ground points are available, the size of the observation vector $y$ will be $2n \times 1$, the size of the design matrix $A_1$ and $A_2$ become $2n \times 6p$ and $2n \times 3m$, respectively. $A_1$ in Eq.(4.35) is formed assuming that the first two image points are in the first photo. Note that $B_{11}$ and $B_{21}$ are in the same column in the upper left part of $A_1$. $A_2$ in Eq.(4.37) is formed assuming that the first and third image points correspond to the first GCP. Note that $C_{11}$ and $C_{31}$ in the same column in the upper left part of $A_2$.

$$ y = A_1 \xi_1 + A_2 \xi_2 + e, \quad e \sim \left( 0, \sigma_\varepsilon^2 P^{-1} \right) $$

(4.33)

Where,
\[
\mathbf{y}^{(2\times n)} = \begin{pmatrix}
- \left( x_1 - x_0 + \Delta x_1 + f \frac{U_1^0}{W_1^0} \right) \\
- \left( y_1 - y_0 + \Delta y_1 + f \frac{V_1^0}{W_1^0} \right) \\
\vdots \\
- \left( x_n - x_0 + \Delta x_n + f \frac{U_n^0}{W_n^0} \right) \\
- \left( y_n - y_0 + \Delta y_n + f \frac{V_n^0}{W_n^0} \right)
\end{pmatrix}
\]

(4.34)

\[
\mathbf{A}_i^{(2\times 6, \rho)} = \begin{bmatrix}
B_{11}^{(2\times 6)} & 0 & 0 & \cdots \\
B_{21}^{(2\times 6)} & 0 & 0 & \cdots \\
0 & B_{32}^{(2\times 6)} & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots \\
\end{bmatrix}
\]

(4.35)

with the sub-matrix for the \(i\)-th image point on the \(j\)-th photo,

\[
\mathbf{B}_{ij}^{(2\times 6)} = \begin{bmatrix}
\frac{\partial F_{x_i}}{\partial \omega_j} & \frac{\partial F_{x_i}}{\partial \phi_j} & \frac{\partial F_{x_i}}{\partial \kappa_j} & \frac{\partial F_{x_i}}{\partial X_{Lj}} & \frac{\partial F_{x_i}}{\partial Y_{Lj}} & \frac{\partial F_{x_i}}{\partial Z_{Lj}} \\
\frac{\partial F_{y_i}}{\partial \omega_j} & \frac{\partial F_{y_i}}{\partial \phi_j} & \frac{\partial F_{y_i}}{\partial \kappa_j} & \frac{\partial F_{y_i}}{\partial X_{Lj}} & \frac{\partial F_{y_i}}{\partial Y_{Lj}} & \frac{\partial F_{y_i}}{\partial Z_{Lj}}
\end{bmatrix}
\]

The partial derivatives with respect to EOP parameters are already shown in Eqs.(4.23) and (4.24).
\[ \xi_{11}^{(6 \times 1)} = \left[ \xi_{11}^{T} \quad \xi_{12}^{T} \quad \cdots \right]^{T} \]

with the sub-vector of the \(-j\)-th photo EOP,

\[
\xi_{1j}^{(6 \times 1)} = \begin{bmatrix} d\omega_j & d\phi_j & d\kappa_j & dX_{Lj} & dY_{Lj} & dZ_{Lj} \end{bmatrix}^{T}
\]

\[ \xi_{2j}^{(2m \times 3m)} = \begin{bmatrix} C_{11}^{(2 \times 3)} & 0 & 0 & \cdots \, 0 \cr 0 & C_{22}^{(2 \times 3)} & 0 & \cdots \, 0 \cr \vdots & \vdots & \vdots & \ddots & \vdots \cr 0 & 0 & 0 & \cdots \, C_{m3}^{(2 \times 3)} \end{bmatrix}
\]

with the submatrix of the \(-i\)-th image point for the \(-j\)-th ground point,

\[
C_{ij}^{(2 \times 3)} = \begin{bmatrix} \frac{\partial F_{x_i}}{\partial X_j} & \frac{\partial F_{x_i}}{\partial Y_j} & \frac{\partial F_{x_i}}{\partial Z_j} \cr \frac{\partial F_{y_i}}{\partial X_j} & \frac{\partial F_{y_i}}{\partial Y_j} & \frac{\partial F_{y_i}}{\partial Z_j} \cr \frac{\partial F_{z_i}}{\partial X_j} & \frac{\partial F_{z_i}}{\partial Y_j} & \frac{\partial F_{z_i}}{\partial Z_j} \end{bmatrix}
\]

\[ \xi_{2j}^{(3 \times 1)} = \begin{bmatrix} \xi_{21}^{(3 \times 1)} & \xi_{22}^{(3 \times 1)} & \cdots \, \xi_{2j}^{(3 \times 1)} \end{bmatrix}^{T} \]

with the sub-vector of the \(-j\)-th photo EOP,

\[
\xi_{2j}^{(3 \times 1)} = \begin{bmatrix} dX_j & dY_j & dZ_j \end{bmatrix}^{T}
\]

The partial derivatives with respect to the GCP coordinate in Eq.(4.37) can be obtained using Eqs.(4.39) and (4.40).
\[
\begin{align*}
\frac{\partial F_x}{\partial X} = & \frac{f(\partial U - U \partial W)}{W}, \\
\frac{\partial F_x}{\partial Y} = & \frac{f(\partial U - U \partial W)}{W}, \\
\frac{\partial F_x}{\partial Z} = & \frac{f(\partial U - U \partial W)}{W}, \\
\frac{\partial F_y}{\partial X} = & \frac{f(\partial V - V \partial W)}{W}, \\
\frac{\partial F_y}{\partial Y} = & \frac{f(\partial V - V \partial W)}{W}, \\
\frac{\partial F_y}{\partial Z} = & \frac{f(\partial V - V \partial W)}{W},
\end{align*}
\] (4.39)

\[
\frac{\partial}{\partial X} \begin{bmatrix} U \\ V \\ W \end{bmatrix} = M \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial}{\partial Y} \begin{bmatrix} U \\ V \\ W \end{bmatrix} = M \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{\partial}{\partial Z} \begin{bmatrix} U \\ V \\ W \end{bmatrix} = M \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\] (4.40)

In addition to linearizing the collinearity equation, the initial values of EOP and GCP are also used as prior information in the observation vector.

First, the linearized EOP observations are obtained as Eq.(4.41). Note that we have \( p \) photos.

\[
\begin{align*}
X_{L_j} = & X_{L_j}^0 + dX_{L_j} + e_{X_{L_j}}, \\
Y_{L_j} = & Y_{L_j}^0 + dY_{L_j} + e_{Y_{L_j}}, \\
Z_{L_j} = & Z_{L_j}^0 + dZ_{L_j} + e_{Z_{L_j}}, \\
\omega_j = & \omega_j^0 + d\omega_j + e_{\omega_j}, \\
\phi_j = & \phi_j^0 + d\phi_j + e_{\phi_j}, \\
\nu = & \nu^0 + d\nu + e_{\nu},
\end{align*}
\] (4.41)

Where,

\[
\begin{bmatrix} e_{EOP_s} \\ \xi \\ e_{EOP_s} \end{bmatrix} \sim \begin{pmatrix} 0, \sigma^2 \end{pmatrix} P_{EOP_s}^{-1}
\]
\[ z_{\text{EOPs}} = \begin{bmatrix} z_{\text{EOPs1}} \vline \ldots \vline z_{\text{EOPs}_p} \end{bmatrix} \]

with sub-vector for the j-th photo, \[ z_{\text{EOPs}_j}^{(6\times1)} = \begin{bmatrix} X_{Lj} - X_{Lj}^0 \\ Y_{Lj} - Y_{Lj}^0 \\ Z_{Lj} - Z_{Lj}^0 \\ \omega_j - \omega_j^0 \\ \phi_j - \phi_j^0 \\ \kappa_j - \kappa_j^0 \end{bmatrix} \] (4.42)

\[ e_{\text{EOPs}} = \begin{bmatrix} e_{\text{EOPs1}} \vline \ldots \vline e_{\text{EOPs}_p} \end{bmatrix} \]

with the sub-vector for the j-th photo, \[ e_{\text{EOPs}_j}^{(6\times1)} = \begin{bmatrix} e_{X_{Lj}} \\ e_{Y_{Lj}} \\ e_{Z_{Lj}} \\ e_{\omega_j} \\ e_{\phi_j} \\ e_{\kappa_j} \end{bmatrix} \] (4.43)

Second, by analogy to GCP, Eq.(4.44) can be formed, which are the linearized GCP observations. Note that we have \( m \) GCP.

\[ \begin{cases} X_j = X_j^0 + dX_j + e_{X_j} \\ Y_j = Y_j^0 + dY_j + e_{Y_j} \Rightarrow z_{\text{GCPs}} \approx I_{(3m\times3m)} \cdot \xi_{(3m\times1)} + e_{\text{GCPs}}, \quad e_{\text{GCPs}} \sim \left(0, \sigma_0^2 P_{\text{GCPs}}^{-1}\right) \end{cases} \] (4.44)

Where, 

\[ \xi = \begin{bmatrix} X \vline Y \vline Z \vline \varphi \vline \theta \vline \psi \end{bmatrix} \]
\[
\begin{align*}
\mathbf{z}_{\text{GCPs}} &= \begin{bmatrix}
\mathbf{z}_{\text{GCPs}1} \\
\vdots \\
\mathbf{z}_{\text{GCPs}m}
\end{bmatrix} \\
\end{align*}
\]  
(4.45)

with the sub-vector for the j-th ground point, \( \mathbf{z}_{\text{GCPs},j} \) 

\[
\begin{align*}
\mathbf{e}_{\text{GCPs}} &= \begin{bmatrix}
\mathbf{e}_{\text{GCPs}1} \\
\vdots \\
\mathbf{e}_{\text{GCPs}m}
\end{bmatrix} \\
\end{align*}
\]  
(4.46)

with the sub-vector for the j-th ground point, \( \mathbf{e}_{\text{GCPs},j} \) 

\[
\begin{align*}
\begin{bmatrix}
X_j - X_j^0 \\
Y_j - Y_j^0 \\
Z_j - Z_j^0
\end{bmatrix}
\end{align*}
\]

Eqs. (4.33), (4.41), and (4.44) form the unified least-squares adjustment, or can be thought of as Gauss Markov model with stochastic constraints (Schaffrin, 2008a), Eq.(4.47).

\[
\begin{align*}
\mathbf{y} &= \mathbf{A}_1 \cdot \hat{\mathbf{x}}_1 + \mathbf{A}_2 \cdot \hat{\mathbf{x}}_2 + \mathbf{e} \\
\mathbf{z}_{\text{EOPs}} &= \mathbf{I} \cdot \hat{\mathbf{x}}_1 + \mathbf{0} \cdot \hat{\mathbf{x}}_2 + \mathbf{e}_{\text{EOPs}} \\
\mathbf{z}_{\text{GCPs}} &= \mathbf{0} \cdot \hat{\mathbf{x}}_1 + \mathbf{I} \cdot \hat{\mathbf{x}}_2 + \mathbf{e}_{\text{GCPs}} \\
\end{align*}
\]  

(4.47)

The estimates of the unknown parameters in the Gauss Markov model with stochastic constraints are computed using Eq.(4.48) that is solved by iterations.
\[
\begin{pmatrix}
\hat{\xi}_1 \\
\hat{\xi}_2
\end{pmatrix} =
\begin{bmatrix}
A_1^T PA_1 + P_{EOPs} & A_1^T PA_2 \\
A_2^T PA_1 & A_2^T PA_2 + P_{GCPs}
\end{bmatrix}^{-1}
\begin{bmatrix}
A_1^T Py + P_{EOPs}z_{EOPs} \\
A_2^T Py + P_{GCPs}z_{GCPs}
\end{bmatrix}
\] (4.48)

The dispersion matrix of the estimated EOP and GCP coordinates is computed from Eq.(4.49).

\[
D_{\xi} \tilde{\begin{pmatrix}
\hat{\xi}_1 \\
\hat{\xi}_2
\end{pmatrix}} = \sigma_0^2 \begin{bmatrix}
A_1^T PA_1 + P_{EOPs} & A_1^T PA_2 \\
A_2^T PA_1 & A_2^T PA_2 + P_{GCPs}
\end{bmatrix}^{-1}
\] (4.49)

Using Eq.(4.47), the residuals can be obtained as shown in Eq.(4.50).

\[
\begin{pmatrix}
\bar{\tilde{\xi}} \\
\tilde{\xi}_{EOPs} \\
\tilde{\xi}_{GCPs}
\end{pmatrix} =
\begin{bmatrix}
A_1 & A_2 \\
I & 0 \\
0 & I
\end{bmatrix}
\begin{pmatrix}
\hat{\xi}_1 \\
\hat{\xi}_2
\end{pmatrix}
\] (4.50)

The estimated variance component is obtained as shown Eq.(4.51).

\[
\hat{\sigma}_0^2 = \bar{\tilde{\xi}} \bar{\tilde{\xi}} \quad 2n
\] (4.51)

Where,

\[n: \text{ the number of image points.}\]
Note that the design matrix in Eq. (4.47) is often large in size, as the number of photos and GCP increase and it requires high computational load, as can be seen in the inverse term in Eq. (4.48). Therefore, it often requires normal equation reduction approach (Mikhail et al., 2000).

4.4.3 Outlier removal (Baarda’s Data Snooping)

In the Baarda’s data snooping method (Baarda, 1968; Schaffrin, 2008a; Schaffrin, 2008b), an outlier will be modeled into the linearized Gauss-Markov (GM) model. To address this, let us begin with GM model with no outlier assumption, shown in Eq. (4.52) as the GM model is the most popular and model for many adjustment problems.

\[ y = A\xi + e, \quad e \sim \left(0, \sigma_0^2 P^{-1}\right) \quad with \ rank(A) = m < n \]  (4.52)

Where,

- \( y \): the observation vector;
- \( A \): the design matrix;
- \( \xi \): the parameter vector;
- \( e \): the random error vector;
- \( n, m \): number of observations, and number of parameters;
- \( \sigma_0^2 \): the variance component (unitless);
- \( P^{-1} \): the cofactor matrix usually given in the form of symmetric and positive-definite.
The least squares estimate is obtained as shown in Eq.(4.53) by minimizing the
Lagrange target function that is the quadratic sum of the random error, \( e^T Pe \) subject to
the condition, \( y - A\xi - e \).

\[
\hat{\xi} = N^{-1}c, \quad \text{with} \quad \hat{\xi} \sim \left( \xi, \sigma_0^2 N^{-1} \right)
\]  
(4.53)

Where,

\[
N = A^T PA, \quad \text{and} \quad c = A^T Py
\]  
(4.54)

The residual vector is computed by substituting the estimate into the GM model as
shown in Eq.(4.55).

\[
\tilde{r} \sim (AN^{-1}A^T) = \sigma_0^2 \tilde{Q}^{-1}
\]  
(4.55)

Where,

\[
\tilde{Q} = \left( I - AN^{-1}A^T \right)
\]  
(4.56)

\( \tilde{Q} \) is cofactor matrix of the residual vector. The estimate of the variance component
is obtained using Eq.(4.57).

\[
\hat{\sigma}_0^2 = \frac{\tilde{r}^2}{n-m} \quad n-m
\]  
(4.57)
with
\[ \Omega = \tilde{\Omega} \quad (4.58) \]

The general case is that the GM model contains outliers \( \xi^{(j)}_0 \) in the \( j \)-th observation.

Note that the GM model is the special case of \( \xi^{(j)}_0 = 0 \) (we can think the GM model is a constrained model with no outlier assumption). Thus, the general observation equation can be expressed as Eq. (4.59).

\[
y = A\xi + \eta_j \xi^{(j)}_0 + e, \quad e \sim \left(0, \sigma_0^2 P^{-1}\right), \quad \text{rank}(A) = m < n \quad (4.59)
\]

Where,
\[
\eta_j = \begin{bmatrix}
0 \\
\vdots \\
0 \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\eta_j \eta_j \\
\end{bmatrix}^T \quad (4.60)
\]

From the Lagrange target function, the normal equations become Eq. (4.61).

\[
\begin{bmatrix}
N & A^T \eta_j \\
\eta_j^T PA & \eta_j^T \eta_j \\
\end{bmatrix}
\begin{bmatrix}
\hat{\xi}^{(j)}_0 \\
\hat{\xi}^{(j)}_0 \\
\end{bmatrix}
= 
\begin{bmatrix}
c \\
\eta_j^T P y \\
\end{bmatrix}
\quad (4.61)
\]

\[ \Rightarrow \hat{\xi}^{(j)}_0 = N^{-1} c - N^{-1} A \eta_j \eta_j \hat{\xi}^{(j)}_0 \quad (4.62) \]
\[ \Rightarrow \hat{\xi}_0^{(j)} = \frac{\eta_j P_j\hat{\xi}}{(PQ)^\top} \]  

(4.63)

\[ \hat{\xi}_0^{(j)} \sim \left( \xi_0^{(j)}, \sigma_0^2 \left( \eta_j PQ \hat{\xi} \right)^{-1} \right) \]  

(4.64)

The residual vector is obtained from Eq.(4.65) using Eqs.(4.59) and (4.62).

\[ \Rightarrow \hat{R}_j \sim \xi_0^{(j)} \]  

(4.65)

The increase in P-weighted quadratic sum of the residual \( R_j \) by introducing the fixed constraint, \( \xi_0^{(j)} = 0 \) is obtained from Eq.(4.66).

\[ R_j = (\hat{\xi}_0^{(j)} - 0) \left( \eta_j PQ \hat{\xi} \right)^{-1} = \eta_j PQ \hat{\xi} \]  

(4.66)

Note that since the GM model is (fixed) constrained with \( \xi_0^{(j)} = 0 \), therefore, the P-weighted quadratic sum of the residuals in GM model \( \Omega \) is represented by Eq.(4.67).

\[ \Omega = \Omega_j + R_j \]  

(4.67)

Therefore, the variance component estimate with j-th observation outlier estimate can be estimated from Eq.(4.68). Note that the number of the unknown parameters increased by one in the denominator.
$$\left( \hat{\sigma}_0^2 \right)^{(j)} = \frac{\Omega_j}{n-(m+1)} = \frac{\Omega - R_j}{n-(m+1)} \quad (4.68)$$

Now, let us consider more simple but practical case, where the observation weight matrix, $P$, is diagonal, i.e. $P = \text{Diag}(p_1, p_2, \ldots)$. Eq.(4.63) becomes Eq.(4.69).

$$\hat{\xi}_0^{(j)} = \frac{\eta_j^T P^w}{(PQ, n)} - (Q, \eta) - \sim \quad (4.69)$$

Where,

$$r_j = (Q, \eta) \quad \text{(4.70)}$$

$r_j$ is called the redundancy number, that can be used to determine how much of the outlier information is reflected in the residuals, as shown in Eq.(4.71).

$$\sim \quad (4.71)$$

Since $(Q_x P)_{gg} = \left( P^{1/2} Q_x P^{1/2} \right)_{gg}$ for diagonal matrix $P$, $r_j$ has the value ranging from -1 to +1. Note that $r_j \in [0, 1]$ for general non-diagonal $P$. The sum of redundancy number is equivalent to the redundancy of the system.
\[ r = n - m = \sum_{j=1}^{n} r_j \]  \hspace{1cm} (4.72)

The increase in the quadratic sum of the residuals Eq.(4.66) is further expanded with Eq.(4.69) as Eq.(4.73).

\[
R_j = \left( \begin{array}{c}
\bar{r}_j \\
\vdots \\
r_j
\end{array} \right) - \bar{r} \cdot \bar{r}^T
\]

\[ \Rightarrow R_j = \left( \begin{array}{c}
\bar{r}_j \\
\vdots \\
r_j
\end{array} \right) \]  \hspace{1cm} (4.73)

Finally, the outlier test procedure can be summarized as follows.

1. Set up hypothesis

“Null hypothesis”: there’s no outlier in \( j \)-th observation vs. “Alternative hypothesis”

\[ H_0 : \xi_0^{(j)} = 0 \quad \text{vs.} \quad H_a : \xi_0^{(j)} \neq 0 \]  \hspace{1cm} (4.74)

2. Compute test statistic

\[
T_j = \frac{R_j / 1}{(\Omega - R_j) / (n - m - 1)} = \frac{R_j (n - m - 1)}{\Omega - R_j}
\]  \hspace{1cm} (4.75)
Ω and \( R_j \) are obtained by Eqs.(4.73), and (4.58).

3. Decision

\[
\begin{align*}
T_j \leq F_\alpha (1, n-m-1) & : \text{Accept } H_0 (\text{no outlier detected}) \\
T_j > F_\alpha (1, n-m-1) & : \text{Reject } H_0 (\text{outlier warning})
\end{align*}
\]  

(4.76)

Where \( F_\alpha \) is F-distribution at \( \alpha \) significance level.

In the bundle adjustment, the unified form of Eq.(4.47) corresponds to Eq.(4.52). Then, subsequent outlier detection process can be carried out.
4.5 Test and Evaluation

4.5.1 Test data

Table 4.3 presents the test data specifications with IKONOS stereo images as reference data and Figure 4.21 and Figure 4.22 show the test area and the aerial images, respectively. Figure 4.21(a) shows the epipolar resampled IKONOS image which is the result of the algorithm introduced in Chapter 3.4. Figure 4.21(b) is a subset image of Figure 4.21(a). A strip of aerial images containing various ground textures including fields, forests and buildings was selected as shown in Figure 4.22. Note that the aerial images are acquired in June 2003 while IKONOS images were collected in November 2001. Due to time and seasonal gap, there are large seasonal and terrain differences between the images. Note that the aerial camera’s interior orientation parameters such as principal point coordinates and radial distortion parameters were known.

As discussed in Chapter 4.2.2, it is important to select a spectral band to be used for image matching, since the aerial image has three spectral bands, while the given IKONOS data are panchromatic (Zhang et al., 2000). In this study, the red band is selected because the IKONOS relative spectral response shows that the red band has the acceptable overlap with panchromatic spectra among the aerial image color bands as depicted in Figure 4.23. In the figure, even though the green band covers a little larger
spectral range, i.e., wider in terms of spectral range, the red band shows better similarity in terms of spectral response.

<table>
<thead>
<tr>
<th>Data</th>
<th>Site</th>
<th>Date</th>
<th>Focal length</th>
<th>Resolution</th>
<th>Spectral</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKONOS, Level2 Stereo</td>
<td>Daejeon, Korea</td>
<td>Nov 2001</td>
<td>10m</td>
<td>1m</td>
<td>Pan</td>
</tr>
<tr>
<td>Aerial images</td>
<td>Daejeon, Korea</td>
<td>Jun 2003</td>
<td>55mm</td>
<td>25cm</td>
<td>Color</td>
</tr>
</tbody>
</table>

Table 4.3 Test data specifications.

Figure 4.21 Test reference images (a) epipolar resampled IKONOS image (b) subset image.
Figure 4.22 Test aerial images, [A][B] show agricultural fields, [D][E] are forests, [G] contains buildings and [C][F] represent mixed areas.
4.5.2 Image matching

First, the initial localization is carried out by coarse matching and the results are presented in Figure 4.24. The first and the third column images are the aerial images with matching points overlaid in them. The second and the fourth columns show the left and the right IKONOS epipolar resampled images with the matching points and initial localization results overlaid in them. Note that aerial images and IKONOS images are Gaussian down-scaled to 2 m spatial resolution for matching. Coarse matching is based on
SIFT + RANSAC using the affine model. The goal of the coarse matching is only localization and the threshold for error rejection in RANSAC is loosely set to $\sigma = 5$ pixels with 99% probability. Figure 4.24 shows that regions of interest could be obtained successfully. In the forest area, image matching results seem acceptable despite the seasonal difference. Since the images are Gaussian down-scaled, the low frequency terrain feature are identified as feature points. Also, another advantage of using stereo images in the matching process can be also identified; e.g., more matching points are obtained than in single image case, by combining the points obtained by matching the aerial photo to the left and right IKONOS images, respectively. In Figure 4.24[B], there is no matching point extracted in the bottom-right corner of the aerial image at the first column which was obtained by matching the aerial image to the left IKONOS. A few matching points at the corner were extracted by matching the aerial image to the right IKONOS image as shown in the third column. It is obvious that one of stereo images should have better similarity in viewing angle than the other.

<table>
<thead>
<tr>
<th>Aerial image</th>
<th>Left IKONOS</th>
<th>Aerial image</th>
<th>Right IKONOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued

Figure 4.24 Coarse matching results based on SIFT + RANSAC with affine equation.
Figure 4.24 continued

<table>
<thead>
<tr>
<th></th>
<th>Aerial image</th>
<th>Left IKONOS</th>
<th>Aerial image</th>
<th>Right IKONOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>[B]</td>
<td><img src="image1" alt="Aerial photo" /></td>
<td><img src="image2" alt="Left IKONOS" /></td>
<td><img src="image3" alt="Aerial photo" /></td>
<td><img src="image4" alt="Right IKONOS" /></td>
</tr>
<tr>
<td>[C]</td>
<td><img src="image5" alt="Aerial photo" /></td>
<td><img src="image6" alt="Left IKONOS" /></td>
<td><img src="image7" alt="Aerial photo" /></td>
<td><img src="image8" alt="Right IKONOS" /></td>
</tr>
<tr>
<td>[D]</td>
<td><img src="image9" alt="Aerial photo" /></td>
<td><img src="image10" alt="Left IKONOS" /></td>
<td><img src="image11" alt="Aerial photo" /></td>
<td><img src="image12" alt="Right IKONOS" /></td>
</tr>
<tr>
<td>[E]</td>
<td><img src="image13" alt="Aerial photo" /></td>
<td><img src="image14" alt="Left IKONOS" /></td>
<td><img src="image15" alt="Aerial photo" /></td>
<td><img src="image16" alt="Right IKONOS" /></td>
</tr>
</tbody>
</table>

continued
Based on the region of interest from the coarse matching, subsets of the IKONOS images with a margin were used for fine matching. Since the spatial resolution of IKONOS and aerial images are 1m and 25cm, respectively, the aerial images were downscaled to 1m for efficient matching. Figure 4.25 shows the fine matching results. The aerial image, and the left and the right IKONOS images are presented together with the matching points. Note that rectangles are the SIFT matching points from the fine matching. Then 3D ground coordinates for each SIFT matched point were computed by IKONOS stereo matching (cross-correlation matching with FFT). Then, the ground coordinates and the corresponding aerial image points were refined by RANSAC with the collinearity equation. The refined matching point locations are marked by triangles in Figure 4.25. In RANSAC, the
threshold for error rejection was set to $\sigma = 1$ pixel with 99% probability. The images [A], [B], [F] and [G] show that acceptable number of matching points could be obtained. However, matching points could hardly be obtained over the forest area due to the lack of unique features and large seasonal differences between aerial and satellite imagery. Unlike the coarse matching which showed moderate image matching performance for the forest area, the fine matching was not successful over the forest area as shown in the aerial images [C], [D] and [E].

![Figure 4.25](image)

Figure 4.25 Fine matching results based on SIFT + RANSAC with collinearity equation.
Figure 4.25 continued

<table>
<thead>
<tr>
<th>Aerial image</th>
<th>Left IKONOS</th>
<th>Right IKONOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>[C]</td>
<td><img src="image" alt="C" /></td>
<td><img src="image" alt="C" /></td>
</tr>
<tr>
<td>[D]</td>
<td><img src="image" alt="D" /></td>
<td><img src="image" alt="D" /></td>
</tr>
<tr>
<td>[E]</td>
<td><img src="image" alt="E" /></td>
<td><img src="image" alt="E" /></td>
</tr>
</tbody>
</table>

continued
Figure 4.25 continued

Figure 4.26 illustrates some samples from Figure 4.25 which show the fine matching results. It is observed that the fine matching could successfully locate identifiable matching points for small and large buildings and road intersection etc, despite the scale difference.
Figure 4.26 Samples of RANSAC-refined matching points.
Figure 4.26 continued
4.5.3 Georeferencing

First, a single photo resection was tested for EOP estimation from the ground control points obtained by the multi-scale image matching. The estimated EOP were compared to the reference EOP known from the bundle adjustment using accurate ground control points and the differences are presented in Table 4.4. As compared to the forest area ([C], [D] and [E]), the open field and the urban (building) areas show better estimation performance. The estimation for image [D] is the worst because as shown in Figure 4.25 there are not enough invariant features in the forest area due to the large seasonal gap between the aerial and IKONOS images. Most aerial images except for the forest areas show EOP positional accuracy better than 20 meters; note that the height was relatively well estimated. Note that East coordinate error is large for [C] and [E] and the pitch angle accuracy is low, accordingly. The low accuracy could be expected from the fine matching result shown in Figure 4.25, as most ground control points are located in the upper part of the image [C] and in the lower part of the image [E]. Since East coordinate is about along image row direction, as shown in Figure 4.21(b), the skewed distribution of points led to a lower estimation accuracy. In terms of attitude estimation performance, a sub-degree accuracy could be achieved in most cases with yaw (heading) angles showing the best accuracy.
### Table 4.4 Error of the estimated EOP using single photo resection.

Table 4.5 presents one standard deviation (1-sigma) precision of the estimated EOP.

The precision for forest area [C] and [E] is roughly two times worse than for the other tested areas, as anticipated. Precision for the forest area [D] is infinity practically and EOP for the area could not be estimated.

### Table 4.5 One standard deviation of the estimated EOP using single photo resection.
Image coordinate residuals for the estimated EOP were computed and presented in Table 4.6. Mean residuals range from three to four pixels. In addition to the residuals, image coordinates computed from the estimated EOP were compared to the coordinates from the reference EOP and the image coordinate errors are presented. Most of mean errors are less than 2.5 meters (about 10 pixels in aerial images) except for the forest areas.

<table>
<thead>
<tr>
<th>Residual and error of aerial image coordinates</th>
<th>Residual [pixels]</th>
<th>Error [pixels]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals</td>
<td>Mean</td>
<td>Max</td>
</tr>
<tr>
<td>[A] Field</td>
<td>2.8</td>
<td>12.4</td>
</tr>
<tr>
<td>[B] Field</td>
<td>2.5</td>
<td>10.5</td>
</tr>
<tr>
<td>[C] Field + Forest</td>
<td>2.9</td>
<td>5.1</td>
</tr>
<tr>
<td>[D] Forest</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>[E] Forest</td>
<td>3.5</td>
<td>14.0</td>
</tr>
<tr>
<td>[F] Forest + building</td>
<td>4.0</td>
<td>11.0</td>
</tr>
<tr>
<td>[G] Building</td>
<td>3.4</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Table 4.6 Image coordinates residuals and errors of the estimated EOP using single photo resection.
Second, a bundle adjustment was performed by generating tie points between adjacent aerial images, taking advantage of image overlap. Table 4.7 lists the EOP estimation errors. Note that the EOP accuracy for the forest area images ([C], [D], and [E]) significantly improved. However, the EOP for some images are a little deteriorated, most likely due to the fact that the EOP errors of the forest area images are distributed to the adjacent images.

<table>
<thead>
<tr>
<th>EOP difference</th>
<th>East [m]</th>
<th>North [m]</th>
<th>Height [m]</th>
<th>Roll [deg]</th>
<th>Pitch [deg]</th>
<th>Yaw [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Field</td>
<td>-2.03</td>
<td>-6.87</td>
<td>-6.81</td>
<td>0.3361</td>
<td>-0.0765</td>
<td>0.1444</td>
</tr>
<tr>
<td>[B] Field</td>
<td>-3.72</td>
<td>-6.38</td>
<td>-6.33</td>
<td>0.3520</td>
<td>-0.1225</td>
<td>0.1250</td>
</tr>
<tr>
<td>[C] Field + Forest</td>
<td>-5.96</td>
<td>-8.79</td>
<td>-5.62</td>
<td>0.4921</td>
<td>-0.1929</td>
<td>0.0777</td>
</tr>
<tr>
<td>[D] Forest</td>
<td>-5.47</td>
<td>-6.91</td>
<td>-4.30</td>
<td>0.4151</td>
<td>-0.1537</td>
<td>0.0063</td>
</tr>
<tr>
<td>[E] Forest</td>
<td>-9.30</td>
<td>-5.64</td>
<td>-2.54</td>
<td>0.3402</td>
<td>-0.3279</td>
<td>-0.0632</td>
</tr>
<tr>
<td>[F] Forest + building</td>
<td>-9.25</td>
<td>-7.10</td>
<td>-0.64</td>
<td>0.3851</td>
<td>-0.3321</td>
<td>-0.0940</td>
</tr>
<tr>
<td>[G] Building</td>
<td>-8.44</td>
<td>-7.56</td>
<td>0.97</td>
<td>0.3745</td>
<td>-0.2943</td>
<td>-0.1198</td>
</tr>
</tbody>
</table>

Table 4.7 Error of the estimated EOP using bundle adjustment (without outlier removal).

As shown in Table 4.8, the overall precision also improved, except for area [A]. Note that the precision is now quite uniform for entire test area. The precision for the height is much better than the precision for the horizontal coordinates, and yaw precision was superior to the precision of other angles.
<table>
<thead>
<tr>
<th>EOP standard deviation</th>
<th>East [m]</th>
<th>North [m]</th>
<th>Height [m]</th>
<th>Roll [deg]</th>
<th>Pitch [deg]</th>
<th>Yaw [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Field</td>
<td>4.64</td>
<td>5.88</td>
<td>1.31</td>
<td>0.2779</td>
<td>0.2158</td>
<td>0.0412</td>
</tr>
<tr>
<td>[B] Field</td>
<td>3.77</td>
<td>5.78</td>
<td>1.41</td>
<td>0.2738</td>
<td>0.1742</td>
<td>0.0362</td>
</tr>
<tr>
<td>[C] Field + Forest</td>
<td>3.31</td>
<td>5.41</td>
<td>1.66</td>
<td>0.2607</td>
<td>0.1518</td>
<td>0.0302</td>
</tr>
<tr>
<td>[D] Forest</td>
<td>2.97</td>
<td>5.37</td>
<td>1.88</td>
<td>0.2613</td>
<td>0.1372</td>
<td>0.0311</td>
</tr>
<tr>
<td>[E] Forest</td>
<td>2.82</td>
<td>5.33</td>
<td>1.87</td>
<td>0.2634</td>
<td>0.1389</td>
<td>0.0336</td>
</tr>
<tr>
<td>[F] Forest + building</td>
<td>3.68</td>
<td>5.59</td>
<td>1.64</td>
<td>0.2784</td>
<td>0.1800</td>
<td>0.0395</td>
</tr>
<tr>
<td>[G] Building</td>
<td>5.72</td>
<td>5.80</td>
<td>1.87</td>
<td>0.2875</td>
<td>0.2776</td>
<td>0.0506</td>
</tr>
</tbody>
</table>

Table 4.8 One standard deviation of the estimated EOP using bundle adjustment (without outlier removal).

Next, a bundle adjustment with outlier removal was carried out using Baarda’s data snooping method (Baarda, 1968). Cumulative F-distribution with $\alpha = 99.99\%$ was used for the hypothesis test as discussed in Chapter 4.4.3. Note that one image point measurement yields two collinearity equations, i.e., one equation for row (line) and one for column (sample) coordinate. Therefore, the image point is flagged as outlier if the null hypothesis is rejected for any one of the two. As depicted in Figure 4.2, the data snooping is iterated until no outlier is detected. The total of six iterations of data snooping were needed to remove all outliers, as shown in Table 4.9. The variance component significantly decreased after the first iteration, and then slowly converged to 0.32. As a result, a total of 20 image points were detected as outliers which corresponds to 12 ground control points.
Table 4.9 Bundle adjustment (Baarda’s data snooping). $F_\alpha$ is F-distribution at $\alpha$ significance level, DOF is degree of freedom.

Table 4.10 presents the outlier detection statistics of several image points for the first bundle adjustment. The image points 92, 94 and 119 were flagged as outliers. Note that the image point 119 was flagged because of the test statistics for $y$ coordinate.

Table 4.10 Flagged outlier image points after the first bundle adjustment.
Figure 4.27 illustrates ground control points that were removed as outliers. Total of 12 ground points were removed which corresponds to 20 outlier image points as discussed above. Note that most of them are from the aerial images [C], [D] and [E] that are the forest areas. Since a small number of matching points were obtained over the forest area, no redundant ground controls were available to successfully refine control points in the matching and RANSAC processes.

Figure 4.27 Removed outlier points distribution on the IKONOS image in the iterative bundle adjustment with data snooping (triangles: retained ground controls, crosses: 12 removed points).
Table 4.11 shows the EOP accuracy for bundle adjustment with outlier removal. The EOP accuracy became more uniform and is better than 7 meters. Note that the flight direction is to the East-direction. The negative bias in the East and North coordinates are due to correlation with the negative errors in the pitch angle and the positive errors in the roll angle, respectively. The building area images ([F] and [G]) show fairly accurate height. High ground height variation in the building area seems to lead to a relatively accurate height estimation. Among the attitude angles, again the yaw angle shows the best accuracy.

<table>
<thead>
<tr>
<th>EOP difference</th>
<th>East [m]</th>
<th>North [m]</th>
<th>Height [m]</th>
<th>Roll [deg]</th>
<th>Pitch [deg]</th>
<th>Yaw [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Field</td>
<td>-6.10</td>
<td>-4.70</td>
<td>-5.05</td>
<td>0.1964</td>
<td>-0.2799</td>
<td>0.0471</td>
</tr>
<tr>
<td>[B] Field</td>
<td>-5.22</td>
<td>-3.65</td>
<td>-3.45</td>
<td>0.1590</td>
<td>-0.2243</td>
<td>0.0397</td>
</tr>
<tr>
<td>[C] Field + Forest</td>
<td>-5.16</td>
<td>-4.68</td>
<td>-2.39</td>
<td>0.2171</td>
<td>-0.2115</td>
<td>0.0343</td>
</tr>
<tr>
<td>[D] Forest</td>
<td>-3.39</td>
<td>-4.51</td>
<td>-1.54</td>
<td>0.2169</td>
<td>-0.1243</td>
<td>0.0217</td>
</tr>
<tr>
<td>[E] Forest</td>
<td>-5.72</td>
<td>-3.21</td>
<td>-0.79</td>
<td>0.1561</td>
<td>-0.2096</td>
<td>-0.0063</td>
</tr>
<tr>
<td>[F] Forest + building</td>
<td>-4.42</td>
<td>-4.23</td>
<td>0.11</td>
<td>0.2023</td>
<td>-0.1309</td>
<td>-0.0198</td>
</tr>
<tr>
<td>[G] Building</td>
<td>-3.76</td>
<td>-4.28</td>
<td>0.53</td>
<td>0.1974</td>
<td>-0.0832</td>
<td>-0.0286</td>
</tr>
</tbody>
</table>

Table 4.11 Error of the estimated EOP using bundle adjustment (with outlier removal).

Table 4.12 shows that precisions of the estimated EOP significantly increased, as compared to the precision presented in Table 4.8. EOP precision is not only high, but also uniform for the test areas. Height and yaw angle still show the best precision.
The ground restitution accuracy for tie points was computed and presented in Table 4.13. Without outlier removal, the bundle adjustment shows horizontal accuracy up to 3.45 meters and vertical accuracy up to 5.77 meters. Considering the existence of outliers, this accuracy is quite good. The effect of the outliers seems to be attenuated by the combined impact of many good ground control and tie points. In contrast, the bundle adjustment with outlier removal improves the horizontal and vertical accuracies to 2.18 and 4.53 meters, respectively. Since 1m resolution stereo images were used as reference, these results are relevant.

The primary error sources affecting the estimation accuracy are satellite image positional accuracy, epipolar image resampling accuracy, image matching between aerial and satellite image, stereo matching, GCP distribution on the aerial image, and image

<table>
<thead>
<tr>
<th>EOP standard deviation</th>
<th>East [m]</th>
<th>North [m]</th>
<th>Height [m]</th>
<th>Roll [deg]</th>
<th>Pitch [deg]</th>
<th>Yaw [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Field</td>
<td>0.94</td>
<td>1.20</td>
<td>0.27</td>
<td>0.0565</td>
<td>0.0436</td>
<td>0.0083</td>
</tr>
<tr>
<td>[B] Field</td>
<td>0.79</td>
<td>1.20</td>
<td>0.30</td>
<td>0.0569</td>
<td>0.0362</td>
<td>0.0073</td>
</tr>
<tr>
<td>[C] Field + Forest</td>
<td>0.68</td>
<td>1.13</td>
<td>0.36</td>
<td>0.0547</td>
<td>0.0310</td>
<td>0.0061</td>
</tr>
<tr>
<td>[D] Forest</td>
<td>0.59</td>
<td>1.13</td>
<td>0.41</td>
<td>0.0548</td>
<td>0.0273</td>
<td>0.0063</td>
</tr>
<tr>
<td>[E] Forest</td>
<td>0.57</td>
<td>1.11</td>
<td>0.40</td>
<td>0.0550</td>
<td>0.0281</td>
<td>0.0067</td>
</tr>
<tr>
<td>[F] Forest + building</td>
<td>0.76</td>
<td>1.16</td>
<td>0.35</td>
<td>0.0578</td>
<td>0.0370</td>
<td>0.0079</td>
</tr>
<tr>
<td>[G] Building</td>
<td>1.15</td>
<td>1.19</td>
<td>0.39</td>
<td>0.0593</td>
<td>0.0560</td>
<td>0.0102</td>
</tr>
</tbody>
</table>

Table 4.12 One standard deviation of the estimated EOP using bundle adjustment (with outlier removal).
overlap. Inaccurate positional information of satellite images will introduce positional error in the extracted ground control information. If the epipolar image resampling is carried out approximately, the process may introduce positional error due to inaccurate image resampling. Note that the developed epipolar image resampling method could achieve near zero image resampling error as shown in Chapter 3.4. Inaccurate image matching between aerial image and stereo HRSI as well as inaccurate stereo image matching will contaminate quality of ground coordinates of matching points. Therefore, robust and accurate image matching techniques are required with the outlier removal process. GCP distribution is also important factor because skewed GCP distribution could lead to inaccurate EOP estimation. Finally, the level of image overlap is also important to carry out bundle adjustment.

<table>
<thead>
<tr>
<th>Ground restitution accuracy</th>
<th>East [m]</th>
<th>North [m]</th>
<th>Height [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bundle adjustment (without outlier removal)</td>
<td>Mean 2.15</td>
<td>0.98</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>Max 3.45</td>
<td>2.69</td>
<td>5.77</td>
</tr>
<tr>
<td>Bundle adjustment (with outlier removal)</td>
<td>Mean 1.15</td>
<td>0.30</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>Max 2.18</td>
<td>0.76</td>
<td>4.53</td>
</tr>
</tbody>
</table>

Table 4.13 Ground restitution accuracy.
4.6 Summary

In this chapter, an automated georeferencing of aerial images using a stereo HRSI pair was proposed. The advantages of using stereo HRSI over the conventional method of using a standard ortho-rectified image and external elevation model can be listed as:

1. A stereo HRSI pair does not require external ground elevation data since the ground elevation can be computed by ground restitution from the stereo.

2. A stereo HRSI pair does not contaminate the quality of the extracted ground control point due to the relief displacement of the ground features, which is present in a standard ortho-rectified image.

3. Aside from high positional accuracy, data update cycle of stereo HRSI is more frequent than conventional reference data such as a ortho-rectified image and ground elevation data.

4. A stereo HRSI pair does not require post geometric correction processing, such as ortho-rectification.

In the proposed approach, a multi-scale image matching scheme was proposed based on SIFT and RANSAC for robustness and efficiency. Using 3D ground coordinates computed from stereo HRSI, the georeferencing was performed using the bundle adjustment with rigorous outlier removal based on Baarda’s data snooping. Major findings from the study can be listed as:
1. The SIFT matching can provide acceptable number of matching points even under varying image conditions such as distortion and spectral differences. Since the SIFT matching accuracy is often low for a large scale mapping, it requires extensive post-processing to secure satisfactory matching accuracy.

2. It is important to select spectral bands when image matching between different types of geospatial imagery is carried out.

3. The multi-scale image matching scheme could successfully extract image points from the aerial images and HRSI.

4. The collinearity condition between 3D ground and image points is useful to prune outlier image points in aerial images by RANSAC.

5. FFT-based normalized cross correlation could significantly reduce the computational cost in the stereo HRSI image matching for 3D ground restitution.

6. Acceptable accuracy of image georeferencing was obtained from a single photo resection over feature-rich targets such as built-up area, but this failed over featureless area such as forested area.

7. The bundle adjustment is required to bridge the ground control gap over those featureless areas, using adjacent images containing good ground features.

8. The iterative bundle adjustment based on the Baarda’s data snooping should be used to detect and remove outliers which may remain after the RANSAC-based outlier removal process.
9. Even though the estimated position and attitude of the platform are not very accurate as explained earlier, the acceptable mapping accuracy can still be obtained.

10. The mapping accuracy is highly dependent on the positional accuracy of stereo HRSI, and the extracted ground control distribution in the aerial images.

11. The estimated position accuracy of the platform is acceptable in the perspective of navigation. Therefore, this approach can be used as a backup for robust navigation under GPS-denied conditions.
5 SUMMARY AND CONCLUSIONS

This study proposed new methods of solving two technical challenges in high-resolution satellite image (HRSI) and aerial image data processing for accurate and robust mapping applications, namely epipolar image resampling of stereo HRSI from a pushbroom sensor, and automated georeferencing of aerial images based on stereo HRSI.

5.1 Epipolar resampling of HRSI

First, this study proposed a new method for epipolar curve determination and epipolar image resampling for spaceborne pushbroom sensors using RPC. Epipolar geometry provides very useful information for processing stereo images. It is crucial not only for significant reduction of image matching cost but also pivotal to stereo-display in the digital photogrammetric workstations by enabling stereo image generation. However, unlike frame cameras that display well-known epipolar geometry, the pushbroom camera does not produce straight epipolar lines and the epipolar pair does not exist for the entire scene. This makes it difficult to exploit the geometry. Still, with a number of high-resolution satellites that are operational, RPC has become the most widely used sensor model, because the users do not need to develop or purchase different sensor models for
different satellite imagery, as RPC provides good accuracy, comparable to the rigorous model. This study proposed a piecewise epipolar curve generation and epipolar resampling method using RPC.

The determination of epipolar curve pair for the entire image was studied, which represents a novel concept. An experiment was carried out that clearly demonstrated that the discrepancy between the unpaired epipolar curves can be negligible for a local image area, which is determined by ground height range of the target. Based on the test result, a piecewise determination of a global epipolar curve pair was proposed by linking together the local epipolar curve pairs. Even though the epipolar curves are not straight, this provided a good basis for establishing the following epipolar image resampling algorithm.

Once epipolar curve pairs were determined for a stereo HRSI pair, a new epipolar resampling algorithm was developed. The basic idea is to arrange the epipolar curve pair points to satisfy the important epipolar image conditions, that is the x-axis of the epipolar resampled image should be aligned along the sensor trajectory while the y-axis should be orthogonal to the trajectory. In addition, there should be no y-parallax and the x-parallax should be linearly proportional to the ground height.

The tests performed on IKONOS stereo images confirmed that the epipolar images could be successfully generated following the RPC update with four GCP. The results showed a maximum y-parallax of 1.3 pixels for manually measured tie points while the conventional method showed a maximum of 4.6 pixels.
The method of piecewise epipolar curve points generation and epipolar image resampling is capable of providing good solutions for accurate 3D topographic mapping and stereo visualization, especially on the digital photogrammetric workstation. In addition to high accuracy, the proposed method has another advantage of compatibility with commercial software which already has the RPC modules, such as Leica Photogrammetry Suite, and SocetSET.

5.2 Satellite Stereo Imagery-based Georeferencing of Aerial Images

The second contribution of this study is the development of a new concept of automatic aerial image georeferencing using stereo HRSI. Significant improvements in HRSI specifications, including high spatial and temporal resolutions, good positional accuracy and large swath width, motivated the idea of using satellite imagery as a reference for image-to-image based indirect georeferencing of airborne imagery. Indirect image georeferencing normally requires accurate 3D ground coordinates. As a new approach, the use of stereo satellite images was proposed here as a reference to provide 3D ground coordinates for georeferencing. Using stereo images can avoid the impact of relief displacement in the reference data and releases the requirement of accurate external height information. In this new approach, a robust and efficient image matching scheme with outlier removal was proposed.
A SIFT-based multi-scale image matching scheme including coarse and fine matching was adopted for efficient matching between aerial and satellite images. The coarse matching performs the initial localization of aerial images in satellite image based on a combination of SIFT and affine model-based RANSAC to model the geometric difference between Gaussian down-scaled images. The coarse matching is followed by the fine image matching to obtain accurate matching points with ground coordinates computed from satellite stereo pairs by cross-correlation matching, implemented by FFT. Then, the collinearity equation-based RANSAC is carried out to refine the matching points.

An experiment was carried out for a strip of aerial images and IKONOS stereo images. The test results demonstrated that good matching results could be obtained over open and built-up areas. However, the matching was poor over forested areas, especially with large seasonal difference between the aerial image and the reference. EOP of the aerial images were estimated from ground control information acquired from the matching points. Both single photo resection and bundle adjustments were tested. Using the bundle adjustment approach with blunder removal, accurate and precise EOP could be obtained and the ground restitution accuracy was at the level that can be expected from the positional accuracy of stereo satellite images. Even though georeferencing performance is highly subjected to many error sources including reference data positional accuracy and matching quality, the high potential of stereo HRSI as reference data has been confirmed. The approach can be used not only for aerial image georeferencing in mapping at a small scale but also provides a good basis for robust navigation under GPS signal loss, since it is
capable of providing position of the platform facilitating image-based terrain-referenced navigation.

The proposed approach is expected to be an alternative for indirect georeferencing that can serve mapping in remote areas as well as terrain-referenced navigation which is gaining momentum as a backup navigation approach in GPS-denied conditions. Since high resolution satellite images are acquired globally with short update cycles, the frequent image update is possible. Moreover, this approach is not only robust to the geometric distortion which often exists on the conventionally used ortho-rectified images, but also does not require external ground height information such as DEM.

5.3 Contributions

Contribution of this doctoral research include:

1. Investigation of the epipolar geometry of RPC, showing that the epipolar curves from RPC have similar properties to those of the rigorous model.

2. Development of the piecewise approach for the epipolar curve pair generation over the entire stereo image pair which cannot be obtained using the conventional epipolar curve generation method.

3. Validation of the piecewise approach by showing that the epipolar curve pair approximately exists for a local area.
4. Development of a new algorithm for accurate epipolar image resampling, and demonstration that high-order polynomial transformation or interpolation should be used than affine model.

5. Development of a new concept of automated aerial image georeferencing using a stereo HRSI.

6. Analysis of the SIFT matching performance for geospatial imagery as a function of varying image parameters and conditions.

7. Development of the multi-scale image matching scheme for robust and efficient image matching between the aerial images and HRSI.

8. Application of the collinearity condition between 3D ground and image points to RANSAC-based outlier image points removal in aerial images for robust image matching between aerial and stereo HRSI.

9. Development of FFT-based normalized cross correlation to significantly reduce the computational cost in the stereo image matching for 3D ground restitution.

10. Application of the iterative bundle adjustment with rigorous outlier removal for reliable aerial image georeferencing.

### 5.4 Future works

In the epipolar resampling of HRSI, the recommendations for future work include performing more investigation of the epipolar geometry of RPC using data over areas with large terrain variations. Also, more experiments on various satellite imagery are
recommended. To generalize the developed algorithm, the direct use of 3D physical sensor model based on the ephemeris, instead of RPC is also recommended for accurate epipolar curve generation to study the curve pattern and the following epipolar image resampling.

The automated aerial image georeferencing using stereo HRSI also requires more experiments on diverse data over different terrain. Also, extensive performance comparisons of the approach will be required compared to the conventional automated georeferencing methods. Future research needs to make RANSAC process more robust since RANSAC is based on random sampling and sometimes produces a different result at each run. It is also recommended to use heterogeneous stereo HRSI pair as a reference data, e.g. IKONOS and Quickbird.
BIBLIOGRAPHY


Schaffrin, B. (2007): Adjustment Computation I, GS650 Lecture Notes, the Ohio State University, Columbus, OH, USA.

Schaffrin, B. (2008a): Adjustment Computation II, GS651 Lecture Notes, the Ohio State University, Columbus, OH, USA.

Schaffrin, B. (2008b): Advanced Adjustment Computation, GS762 Lecture Notes, the Ohio State University, Columbus, OH, USA.


APPENDIX A: RPC MATHEMATICAL MODEL

The basic RFM equation in the forward projection form, i.e., a ground coordinate to the image point coordinate, is expressed as Eq.(A.1). For the given ground point coordinates, \((\phi, \lambda, h)\), the corresponding image coordinates \((l,s)\) can be computed. The equation is a nonlinear equation of 80 coefficients (RPCs), i.e. \(a, b, c, d\).

\[
Y = \frac{a^T u}{b^T u}, \quad X = \frac{c^T u}{d^T u}
\] (A.1)

with

\[
U = \frac{\phi - \phi_0}{\phi_s}, \quad V = \frac{\lambda - \lambda_0}{\lambda_s}, \quad W = \frac{h - h_0}{h_s}, \quad Y = \frac{l - L_0}{L_s}, \quad X = \frac{s - S_0}{S_s}
\] (A.2)

\[
u = \begin{bmatrix}
1 & V & U & W & UV & VW & UW & V^2 \\
U^2 & W^2 & UVW & V^3 & VU^2 & VW^2 & V^2U \\
U^3 & UW^2 & V^2W & U^2W & W^3
\end{bmatrix}^T
\] (A.3)
\[ a = [a_1 \ a_2 \ \ldots] \quad (A.4) \]
\[ b = [1 \ b_2 \ \ldots] \quad (A.5) \]
\[ c = [c_1 \ c_2 \ \ldots] \quad (A.6) \]
\[ d = [1 \ d_2 \ \ldots] \quad (A.7) \]

Where,

\[ X, Y \] the normalized image space coordinates of ground target;
\[ U, V, W \] the normalized object space coordinates of ground target;
\[ \phi, \lambda, h \] the geodetic latitude, longitude and ellipsoidal height of ground target;
\[ l, s \] the image line (row) and sample (column) coordinates;
\[ \phi_0, \lambda_0, h_0, S_0, L_0 \] the offset factors for the latitude, longitude, height, sample and line;
\[ \phi_s, \lambda_s, h_s, S_s, L_s \] the scale factors for the latitude, longitude, height, sample and line.

A.1. Image to ground projection (to the fixed ground height)

Given an image point \((l, s)\) and given height \(h\), the horizontal ground coordinate can be computed from RFM. The given information should be normalized first as Eq.(A.8).

\[ Y = \frac{l - L_0}{L_s}, \quad X = \frac{s - S_0}{S_s}, \quad W = \frac{h - h_0}{h_s} \quad (A.8) \]
Since the RFM is not in linear form, Taylor series expansion is applied first to obtain the linearized RFM equation. Note that the initial horizontal ground coordinates are required for the expansion and it can be obtained from the first order terms by neglecting higher order terms in RFM as Eq.(A.9) which is organized for Eq.(A.10).

\[
Y = \frac{a_1 + a_2 V^0 + a_3 U^0 + a_4 W}{b_1 + b_2 V^0 + b_3 U^0 + b_4 W} \\
X = \frac{c_1 + c_2 V^0 + c_3 U^0 + c_4 W}{d_1 + d_2 V^0 + d_3 U^0 + d_4 W}
\]  

(A.9)

\[
\begin{bmatrix}
Y^0 \\
U^0
\end{bmatrix}
= \begin{bmatrix}
(Yb_2 - a_2) & (Yb_3 - a_3) \\
(Xd_2 - c_2) & (Xd_3 - c_3)
\end{bmatrix}^{-1}
\begin{bmatrix}
a_1 + (a_4 - Yb_4)W - Yb_1 \\
c_1 + (c_4 - Xd_4)W - Xd_1
\end{bmatrix}
\]  

(A.10)

Now that the initial normalized horizontal ground coordinates \((V^0, U^0)\) are computed, the linearization of RFM (Eq.(A.1)) is obtained with respect to the normalized horizontal ground coordinate \((V, U)\) as Eq.(A.11).

\[
Y = Y^0 + \frac{\partial Y^0}{\partial V} dV + \frac{\partial Y^0}{\partial U} dU \\
X = X^0 + \frac{\partial X^0}{\partial V} dV + \frac{\partial X^0}{\partial U} dU
\]  

(A.11)

Where,
\[
Y^0 = \frac{a^T u^0}{b^T u^0}, \quad X^0 = \frac{c^T u^0}{d^T u^0} \quad (A.12)
\]

\[
u^0 = \left[1 \quad V^0 \quad U^0 \quad W \quad V^0 U^0 \quad V^0 W \quad U^0 W \quad (V^0)^2 \quad (U^0)^2 \quad V^0 (U^0)^2 \quad V^0 W^2 \quad (V^0)^2 U^0 \quad (U^0)^3 \quad V^0 (U^0)^2 \quad V^0 W^2 \quad (V^0)^2 U^0 \quad W^3 \right]^T \quad (A.13)
\]

The coefficients in Eq. (A.11) are obtained by the chain rule for derivatives as below:

\[
\frac{\partial Y}{\partial V} = \frac{\partial Y}{\partial u} \frac{\partial u}{\partial V} \quad (A.14)
\]

\[
\frac{\partial Y}{\partial u} = \frac{\partial Y}{\partial u} \quad (A.15)
\]

\[
\frac{\partial X}{\partial V} = \frac{\partial X}{\partial u} \quad (A.16)
\]

\[
\frac{\partial X}{\partial u} = \frac{\partial X}{\partial u} \quad (A.17)
\]

with

\[
\frac{\partial \left( a^T u \right)}{\partial u} = \left( b^T u \right)^2 \frac{\partial \left( b^T u \right)}{\partial u} - \left( a^T u \right) \frac{\partial \left( b^T u \right)}{\partial u} = \frac{\partial \left( b^T u \right)}{\partial u} \left( b^T u \right)^2 - \left( a^T u \right) b^T \quad (A.18)
\]

\[
\frac{\partial \left( c^T u \right)}{\partial u} = \left( d^T u \right)^2 \frac{\partial \left( d^T u \right)}{\partial u} - \left( c^T u \right) \frac{\partial \left( d^T u \right)}{\partial u} = \frac{\partial \left( d^T u \right)}{\partial u} \left( d^T u \right)^2 - \left( c^T u \right) d^T \quad (A.19)
\]

\[
\frac{\partial u}{\partial U} = \begin{bmatrix} 0 & 0 & 1 & 0 & V & 0 & W & 0 & 2U & 0 & V W & 0 & 2V U & 0 & V^2 & 3U^2 & W^2 & 0 & 2U W & 0 \end{bmatrix}^T \quad (A.20)
\]
\[
\begin{align*}
\frac{\partial u}{\partial V} &= \begin{bmatrix} 0 & 1 & 0 & 0 & U & W & 0 & 2V & 0 & 0 & UW & 3V^2 & U^2 & 2VU & 0 & 0 & 2W & 0 & 0 \end{bmatrix}^T \quad (A.21) \\
\frac{\partial u}{\partial W} &= \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & V & U & 0 & 0 & 2W & UV & 0 & 0 & 2VW & 0 & 0 & 2UW & V^2 & U^2 & 3W^2 \end{bmatrix}^T \quad (A.22)
\end{align*}
\]

The incremental coordinates can be solved by Eq. (A.23) and the normalized horizontal ground coordinates are updated as Eq. (A.24). The normalized horizontal ground coordinates can be computed through iterative estimation process where the incremental change is ignorable.

\[
\begin{bmatrix} dV \\ dU \end{bmatrix} = \begin{bmatrix} \frac{\partial Y}{\partial V} & \frac{\partial Y}{\partial U} \\ \frac{\partial X}{\partial V} & \frac{\partial X}{\partial U} \end{bmatrix}_{V=V^0, U=U^0}^{-1} \begin{bmatrix} Y - Y^0 \\ X - X^0 \end{bmatrix} \quad (A.23)
\]

\[
\begin{bmatrix} V \\ U \end{bmatrix} = \begin{bmatrix} V^0 \\ U^0 \end{bmatrix} + \begin{bmatrix} dV \\ dU \end{bmatrix} \quad (A.24)
\]

Finally, the horizontal ground coordinate, \((\varphi, \lambda)\), is obtained using Eq. (A.2).

**A.2. Space intersection (Ground restitution)**

To restitute the 3D ground coordinates from a pair of image points of stereo images, Taylor series expansion is applied to RFM to obtain the linearized equation with respect to the ground coordinates. Note that the initial 3D ground coordinates are required for the
expansion, and it can be obtained from the first order terms. For a single image \((i)\), Eq.(A.25) can be obtained by neglecting higher order terms and derived further for the 3D ground coordinates as Eq.(A.26).

\[
Y^{(i)} = \frac{a_1^{(i)} + a_2^{(i)}v^0 + a_3^{(i)}u^0 + a_4^{(i)}w^0}{b_1^{(i)} + b_2^{(i)}v^0 + b_3^{(i)}u^0 + b_4^{(i)}w^0} \\
X^{(i)} = \frac{c_1^{(i)} + c_2^{(i)}v^0 + c_3^{(i)}u^0 + c_4^{(i)}w^0}{d_1^{(i)} + d_2^{(i)}v^0 + d_3^{(i)}u^0 + d_4^{(i)}w^0}
\]

(A.25)

Where \((i)\) is image number

\[
\begin{align*}
Y^{(i)}b_1^{(i)} + Y^{(i)}b_2^{(i)}v^0 + Y^{(i)}b_3^{(i)}u^0 + Y^{(i)}b_4^{(i)}w^0 &= a_1^{(i)} + a_2^{(i)}v^0 + a_3^{(i)}u^0 + a_4^{(i)}w^0 \\
X^{(i)}d_1^{(i)} + X^{(i)}d_2^{(i)}v^0 + X^{(i)}d_3^{(i)}u^0 + X^{(i)}d_4^{(i)}w^0 &= c_1^{(i)} + c_2^{(i)}v^0 + c_3^{(i)}u^0 + c_4^{(i)}w^0
\end{align*}
\]

\[
\begin{align*}
\Rightarrow \begin{cases} 
Y^{(i)}b_2^{(i)} - a_2^{(i)}v^0 + (Y^{(i)}b_3^{(i)} - a_3^{(i)})u^0 + (Y^{(i)}b_4^{(i)} - a_4^{(i)})w^0 &= a_1^{(i)} - Y^{(i)}b_1^{(i)} \\
X^{(i)}d_2^{(i)} - c_2^{(i)}v^0 + (X^{(i)}d_3^{(i)} - c_3^{(i)})u^0 + (X^{(i)}d_4^{(i)} - c_4^{(i)})w^0 &= c_1^{(i)} - X^{(i)}d_1^{(i)}
\end{cases}
\]

\[
\Rightarrow \begin{bmatrix}
a_1^{(i)} - Y^{(i)}b_1^{(i)} \\
\end{bmatrix} = \begin{bmatrix}
Y^{(i)}b_2^{(i)} - a_2^{(i)} \\
Y^{(i)}b_3^{(i)} - a_3^{(i)} \\
Y^{(i)}b_4^{(i)} - a_4^{(i)}
\end{bmatrix} \begin{bmatrix}
v^0 \\
u^0 \\
w^0
\end{bmatrix} = \begin{bmatrix}
\end{bmatrix}
\]

(A.26)

Given two images \((i)\) and \((i+1)\), the initial normalized 3D ground coordinates can be computed by Gauss Markov model as Eq.(A.27).

\[
y = A_{(4\times1)} \xi + e_{(4\times1)}
\]

(A.27)
Where,

\[
y^{(4x1)} = \begin{bmatrix}
a^{(i)}_1 - Y^{(i)}b^{(i)}_1 \\
c^{(i)}_1 - X^{(i)}d^{(i)}_1 \\
a^{(i+1)}_1 - Y^{(i+1)}b^{(i+1)}_1 \\
c^{(i+1)}_1 - X^{(i+1)}d^{(i+1)}_1
\end{bmatrix}
\]  
(A.28)

\[
A^{(4x3)} = \begin{bmatrix}
(Y^{(i)}b^{(i)}_2 - a^{(i)}_2) & (Y^{(i)}b^{(i)}_3 - a^{(i)}_3) & (Y^{(i)}b^{(i)}_4 - a^{(i)}_4) \\
(X^{(i)}d^{(i)}_2 - c^{(i)}_2) & (X^{(i)}d^{(i)}_3 - c^{(i)}_3) & (X^{(i)}d^{(i)}_4 - c^{(i)}_4) \\
(Y^{(i+1)}b^{(i+1)}_2 - a^{(i+1)}_2) & (Y^{(i+1)}b^{(i+1)}_3 - a^{(i+1)}_3) & (Y^{(i+1)}b^{(i+1)}_4 - a^{(i+1)}_4) \\
(X^{(i+1)}d^{(i+1)}_2 - c^{(i+1)}_2) & (X^{(i+1)}d^{(i+1)}_3 - c^{(i+1)}_3) & (X^{(i+1)}d^{(i+1)}_4 - c^{(i+1)}_4)
\end{bmatrix}
\]  
(A.29)

\[
\xi^{(3x1)} = \begin{bmatrix}
V^0 \\
U^0 \\
W^0
\end{bmatrix}
\]  
(A.30)

The estimate of the initial normalized 3D ground coordinates are obtained by the least-squares solution as Eq.(A.31).

\[
\hat{\xi} = (A^T PA)^{-1} A^T Py
\]  
(A.31)

Where \( P \) is weight matrix that can be set to the identity matrix.

Following the initial value computation, the update equation Eq.(A.32) is given by linearization of RFM with respect to the normalized 3D ground coordinates \( U, V, W \).
\[ Y^{(i)} = Y^{(i)0} + \frac{\partial Y^{(i)}}{\partial V} \left[ dV + \frac{\partial Y^{(i)}}{\partial U} \right] dU + \frac{\partial Y^{(i)}}{\partial W} dW \]

\[ X^{(i)} = X^{(i)0} + \frac{\partial X^{(i)}}{\partial V} \left[ dV + \frac{\partial X^{(i)}}{\partial U} \right] dU + \frac{\partial X^{(i)}}{\partial W} dW \]  

(A.32)

Where,

\[ Y^{(i)0} = \frac{a^{(i)T} u^0}{b^{(i)T} u^0}, \quad \text{and} \quad X^{(i)0} = \frac{c^{(i)T} u^0}{d^{(i)T} u^0} \]  

(A.33)

\[ u^0 = \begin{bmatrix} 1 & V^0 & U^0 & W & V^0 U^0 & V^0 W & U^0 W & (V^0)^2 & U^0 V^0 W & (V^0)^3 & V^0 (U^0)^2 & V^0 (W^0)^2 & (V^0)^2 U^0 & (U^0)^2 & U^0 (W^0)^2 & (V^0)^2 W & (U^0)^2 W & (W^0)^3 \end{bmatrix} \]  

(A.34)

The coefficients in Eq. (A.32) are obtained by the chain rule for derivatives as below:

\[ \frac{\partial Y^{(i)}}{\partial V} = \frac{\partial Y^{(i)}}{\partial u^T} \frac{\partial u}{\partial V} \]  

(A.35)

\[ \frac{\partial Y^{(i)}}{\partial U} = \frac{\partial Y^{(i)}}{\partial u^T} \frac{\partial u}{\partial U} \]  

(A.36)

\[ \frac{\partial Y^{(i)}}{\partial W} = \frac{\partial Y^{(i)}}{\partial u^T} \frac{\partial u}{\partial W} \]  

(A.37)

\[ \frac{\partial X^{(i)}}{\partial V} = \frac{\partial X^{(i)}}{\partial u^T} \frac{\partial u}{\partial V} \]  

(A.38)

\[ \frac{\partial X^{(i)}}{\partial U} = \frac{\partial X^{(i)}}{\partial u^T} \frac{\partial u}{\partial U} \]  

(A.39)
with
\[
\frac{\partial X^{(i)}}{\partial u^T} = \frac{\partial}{\partial u^T} \frac{\partial X^{(i)}}{\partial u} = \frac{\partial (a^{(i)^T} u) - (a^{(i)^T} u) \partial (b^{(i)^T} u)}{(b^{(i)^T} u)^2} = \frac{(b^{(i)^T} u) a^{(i)^T} - (a^{(i)^T} u) b^{(i)^T}}{(b^{(i)^T} u)^2} \tag{A.41}
\]

\[
\frac{\partial X^{(i)}}{\partial u^T} = \frac{\partial (c^{(i)^T} u) - (c^{(i)^T} u) \partial (d^{(i)^T} u)}{(d^{(i)^T} u)^2} = \frac{(d^{(i)^T} u) c^{(i)^T} - (c^{(i)^T} u) d^{(i)^T}}{(d^{(i)^T} u)^2} \tag{A.42}
\]

\[
\frac{\partial u}{\partial U} = \begin{bmatrix} 0 & 0 & 0 & V & 0 & W & 0 & 2U & 0 & VV & 0 & WU & 0 & 2UW & 0 & V^2 & 2V^2 & W^2 & 0 & 3U^2 & W^2 & 0 & 2UW & 0 & U^2 & 3V^2 & W^2 & 0 & 2UW & 0 & U^2 & 3V^2 & W^2 & 0 & 2UW & 0 & U^2 & 3V^2 & W^2 & 0 \end{bmatrix}^T \tag{A.43}
\]

\[
\frac{\partial u}{\partial V} = \begin{bmatrix} 0 & 1 & 0 & 0 & U & W & 0 & 2V & 0 & 0 & UW & 3V^2 & W^2 & 2VU & 0 & 0 & 2VW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 & 0 & 2UW & 0 \end{bmatrix}^T \tag{A.44}
\]

\[
\frac{\partial u}{\partial W} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & V & U & 0 & 0 & 2W & UV & 0 & 0 & 2W & 0 & 0 & 2W & 0 & 0 & 2W & 0 & 0 & 2W & 0 & 0 & 2W & 0 & 0 & 2W & 0 & 0 & 2W & 0 \end{bmatrix}^T \tag{A.45}
\]

Given two images \((i)\) and \((i+1)\), the update of the normalized 3D ground coordinates can be computed by Gauss Markov model as Eq.(A.46).

\[
y = A \begin{bmatrix} (4 \times 1) \\ (4 \times 3) \\ (3 \times 1) \\ (4 \times 1) \end{bmatrix} \xi + e \tag{A.46}
\]

Where,
The estimates of the normalized 3D ground coordinates update are obtained as Eq.(A.50), and the update equations are shown in Eq.(A.51).

\[
\dot{\xi} = \left( A^T P A \right)^{-1} A^T P y
\]

\[\text{(A.50)}\]

Where \( P \) is weight matrix and can be set to be the identity matrix.
\[
\begin{bmatrix}
V
U
W
\end{bmatrix} = \begin{bmatrix}
V^0
U^0
W^0
\end{bmatrix} + \begin{bmatrix}
dV
dU
dW
\end{bmatrix}
\]  (A.51)

### A.3. RPC Resection

The RPC resection is to indirectly compute RPC from the well established GCP, i.e. estimation \(a, b, c, d\) with \(u, X, Y\) given. In the following, the condensed version of the algorithm is be presented. For a detail derivation, refer to Tao and Yong (2001). First, RFM can be expressed as Eq.(A.52), by differencing the left side of RFM from the right side.

\[
e_r = \frac{a^\top u}{b^\top u} - Y
\]
\[
e_x = \frac{c^\top u}{d^\top u} - X
\]  (A.52)

Where \(e\) is random error.

By letting the denominators \(b^\top u = B\) and \(d^\top u = D\), Eq.(A.52) can be expressed as Eq.(A.53).
\[ e_\gamma = \begin{bmatrix} \frac{u^T}{B} & -Y \frac{u^T}{B} \\ & \begin{bmatrix} a \\ b \end{bmatrix} \end{bmatrix} \]
\[ e_\chi = \begin{bmatrix} \frac{u^T}{D} & -X \frac{u^T}{D} \\ & \begin{bmatrix} c \\ d \end{bmatrix} \end{bmatrix} \]

(A.53)

Where,

\[ B = b^T u \]
\[ D = d^T u \]

(A.54)

By removing the first element from \( u, b, d \), new variables \( u', b', d' \) are defined as Eqs.(A.55) and (A.56). Then Eq.(A.53) becomes Eq.(A.58) via Eq.(A.57).

\[ u' = \begin{bmatrix} V & U & W & VU & VW & UW & V^2 & U^2 & W^2 & UVW & V^3 & VU^2 & VW^2 & V^2U & U^3 & UW^2 & V^2W & U^2W & W^3 \end{bmatrix}^T \]

(A.55)

\[ b' = \begin{bmatrix} b_2 & b_3 & \ldots \end{bmatrix} \]

(A.56)

\[ d' = \begin{bmatrix} d_2 & d_3 & \ldots \end{bmatrix} \]

\[ e_\gamma = \begin{bmatrix} \frac{u^T}{B} & -Y \frac{u^T}{B} \\ & \begin{bmatrix} a \\ b' \end{bmatrix} - \frac{Y}{B} \end{bmatrix} \]
\[ e_\chi = \begin{bmatrix} \frac{u^T}{D} & -X \frac{u^T}{D} \\ & \begin{bmatrix} c \\ d' \end{bmatrix} - \frac{X}{D} \end{bmatrix} \]

(A.57)
\[
\frac{Y}{B} = \begin{bmatrix} u^T \frac{B}{Y} \end{bmatrix} - \begin{bmatrix} u^T \frac{B}{X} \end{bmatrix} + e_y
\]

\[
\frac{X}{D} = \begin{bmatrix} u^T \frac{D}{X} \end{bmatrix} - \begin{bmatrix} u^T \frac{D}{Y} \end{bmatrix} + e_x
\]  

(A.58)

Given \( n \) GCP, Eq.(A.58) can be formed into Eq.(A.59).

\[
\begin{bmatrix}
\frac{Y_1}{B_1} \\
\vdots \\
\frac{Y_n}{B_n}
\end{bmatrix} = \begin{bmatrix}
\begin{bmatrix} u^T \frac{B_1}{B_1} & -Y_1 \frac{u^T \frac{B_1}{B_1}}{B_1} & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & u^T \frac{B_n}{B_n} & -Y_n \frac{u^T \frac{B_n}{B_n}}{B_n} & 0 & 0 & \end{bmatrix}
\end{bmatrix} + \begin{bmatrix}
\begin{bmatrix} e_{y_1} \\
\vdots \\
e_{y_n}
\end{bmatrix}
\end{bmatrix}
\]

(A.59)

Eq.(A.59) is expressed in the Gauss Markov model as Eq.(A.60).

\[
\mathbf{y}_{(2n \times 1)} = \mathbf{A}_{(2n \times 78)} \mathbf{\xi}_{(78 \times 1)} + \mathbf{e}_{(2n \times 1)}
\]  

(A.60)

Where,

\[
\mathbf{y}_{(2n \times 1)} = \begin{bmatrix}
\frac{Y_1}{B_1} \\
\vdots \\
\frac{Y_n}{B_n}
\end{bmatrix}
\]

(A.61)
The solution of Eq. (A.65) can be obtained from iterative least-squares by setting the initial values $\xi^0$, e.g. all zero, in the design matrix $A$. Note that the zero crossing check should be checked in the denominator $B$ and $D$ or every iteration to avoid the division with zero, such as in Eq. (A.59).

$$\hat{\xi} = \left( A^T P A \right)^{-1} A^T P y$$

(A.65)

Where $P$ is weight matrix that be set to an identity matrix.

Tao and Yong (2001) showed that the design matrix $A$ is ill-conditioned and the solution does not converge due to the uneven distribution of the control points. The author suggested the use of error variance to regularize the normal matrix as shown in
Eq.(A.66) and a value of 0.0002 to 0.004 suggested as $h^2$ to achieve numerical convergence.

$$\hat{\xi} = \left( A^T P A + h^2 I_{(78 \times 78)} \right)^{-1} A^T P y$$  \hspace{1cm} (A.66)