

DIGITAL SURFACE MODELS FROM SPACEBORNE IMAGES WITHOUT
GROUND CONTROL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YOLDAŞ ATASEVEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

SEPTEMBER 2012

Approval of the thesis:

**DIGITAL SURFACE MODELS FROM SPACEBORNE IMAGES
WITHOUT GROUND CONTROL**

submitted by **YOLDAŞ ATASEVEN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN

Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet Erkmen

Head of Department, **Electrical and Electronics Engineering**

Prof. Dr. A. Aydın Alatan

Supervisor, **Electrical and Electronics Engineering Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Uğur GÜDÜKBAY

Computer Engineering Dept., Bilkent University

Prof. Dr. A. Aydın Alatan

Electrical and Electronics Engineering Dept., METU

Prof. Dr. Uğur HALICI

Electrical and Electronics Engineering Dept., METU

Prof. Dr. Şebnem DÜZGÜN

Mine Engineering Dept., METU

Prof. Dr. Gözde BOZDAĞI AKAR

Electrical and Electronics Engineering Dept., METU

Date: 18 September 2012

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required, I have fully cited and referenced all material and results that are not original to this work.

Name Lastname : Yoldaş Ataseven

Signature :

ABSTRACT

DIGITAL SURFACE MODELS FROM SPACEBORNE IMAGES WITHOUT GROUND CONTROL

Ataseven, Yoldaş

Ph.D., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. A. Aydın Alatan

September 2012, 229 pages

Generation of Digital Surface Models (DSMs) from stereo satellite (spaceborne) images is classically performed by Ground Control Points (GCPs) which require site visits and precise measurement equipment. However, collection of GCPs is not always possible and such requirement limits the usage of spaceborne imagery. This study aims at developing a fast, fully automatic, GCP-free workflow for DSM generation. The problems caused by GCP-free workflow are overcome using freely-available, low-resolution static DSMs (LR-DSM). LR-DSM is registered to the reference satellite image and the registered LR-DSM is used for *i*) correspondence generation and *ii*) initial estimate generation for 3-D reconstruction. Novel methods are developed for bias removal for LR-DSM registration and bias equalization for projection functions of satellite imaging. The LR-DSM

registration is also shown to be useful for computing the parameters of simple, piecewise empirical projective models. Recent computer vision approaches on stereo correspondence generation and dense depth estimation are tested and adopted for spaceborne DSM generation.

The study also presents a complete, fully automatic scheme for GCP-free DSM generation and demonstrates that GCP-free DSM generation is possible and can be performed in much faster time on computers. The resulting DSM can be used in various remote sensing applications including building extraction, disaster monitoring and change detection.

Key words: Digital Surface Model, GCP, SRTM, Satellite, Spaceborne, Earth Observation, Photogrammetry.

ÖZ

UYDU GÖRÜNTÜLERİNDEN YER KONTROL NOKTASIZ SAYISAL YÜZEY HARİTALARI

Ataseven, Yoldaş

Doktora, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. A. Aydın Alatan

Eylül 2012, 229 sayfa

Uydu görüntülerinden Sayısal Yüzey Modelleri (SYM) oluşturma; genellikle Yer Kontrol Noktaları (YKN) kullanılarak gerçekleştirilir. YKN elde edilmesi, saha ziyaretleri ve hassas ölçüm araçlarının kullanımını gerektirir. Ancak, YKN toplama işlemi her zaman mümkün olmamakta, YKN'lere bağımlılık uydu görüntülerinin kullanılabilirliğini sınırlamaktadır. Bu çalışmada; YKN gerektirmeyen, tam otomatik ve hızlı bir SYM oluşturma yöntemi geliştirilmesi amaçlanmıştır. YKN yoksunluğundan kaynaklanan sorunlar, ücretsiz düşük çözünürlüklü SYM'ler (DÇ-SYM) kullanılarak azaltılmıştır. DÇ-SYM uydu görüntüleriyle hizalanmış; hizalanmış DÇ-SYM kullanılarak; *i*) stereo nokta çiftleri oluşturma ve *ii*) üç boyutlu geriçatımda ilk tahmin oluşturma işlemleri gerçekleştirilmiştir. DÇ-SYM hizalamasında ve uydu görüntüleme model parametrelerindeki kaymaların YKN

olmaksızın giderilmesi için özel yöntemler geliştirilmiştir. Hizalanmış DÇ-SYM'lerin basitleştirilmiş parçalı projeksiyon modelleri için parametre kestiriminde kullanılabileceği gösterilmiştir. Yakın dönemde geliştirilen bilgisayarla görme yöntemleri, uydu görüntülerine dayalı SYM oluşturulmasında kullanılmış, uydu görüntülemesine uygun, melez bir yöntem geliştirilmiştir.

Bu çalışmada; YKN olmaksızın SYM oluşturan, tam otomatik ve bütünsel bir yöntem geliştirilmiş ve uygulanmış; YKN kullanılmadan SYM oluşturmanın mümkün olduğu ve bunun ortalama bilgisayarlar üzerinde kısa sürede gerçekleştirilebileceği gösterilmiştir. Üretilen SYM'ler; bina tespiti, doğal afet izleme, ve fark tespitine yönelik uzaktan algılama çalışmalarında girdi olarak kullanılabilir niteliktedir.

Anahtar kelimeler: Sayısal Yüzey Modeli, Yer Kontrol Noktası, SRTM, Uydu, Yeryüzü Gözlemi.

To my parents and sister

ACKNOWLEDGEMENTS

The author wishes to express his deepest gratitude to his supervisor, Prof. Dr. Aydın Alatan for his keen and intelligent guidance, and insight throughout the research. His wonderful personality, patience and thoughtfulness made this study possible.

The author also wishes to thank Prof.Dr. Şebnem Düzgün, Dr. Ali Özgün Ok and Ms. Hilal Soydan for the data they provided and patient explanations on remote sensing and satellite imaging.

Dear friends Dr. Cevahir Çıgla and Emrah Taşlı have contributed to the study in dense DSM generation with code and explanations. The author is thankful for all.

Another dear friend, Kutalmış Gökalp İnce helped the author with brilliant ideas and intelligent comments. The idea of bias equalization is triggered by him.

A study on satellite imaging requires data, which are expensive. Thanks to the freely available data provided by JAXA, NASA and ISPRS, this study became possible.

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), with National Scholarship Programme for Ph.D. Students.

TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZ.....	vi
ACKNOWLEDGEMENTS.....	ix
TABLE OF CONTENTS	x
CHAPTER	
1 INTRODUCTION	1
1.1 General Overview.....	1
1.2 The Addressed Reader.....	2
1.3 Problem Statement	3
1.4 Error Measures.....	4
1.5 Satellite Imagery	6
1.5.1 The Imaging Scenario	7
1.5.2 The Collinearity Equations	12
1.5.3 The Coplanarity Constraints	14
1.5.4 Satellite Imaging Systems	15
1.5.5 Coordinate Systems	18
1.5.6 Earth Models.....	21
1.6 Data Types.....	22
1.6.1 Monochrome / Panchromatic Images.....	22
1.6.2 Multispectral / Colored Images.....	23
1.6.3 Hyperspectral Images.....	24
1.6.4 Synthetic Aperture Radar	25
1.6.5 Ground Control Points.....	25
1.6.6 Low Resolution Digital Surface Models	26
1.7 Sources of Satellite Data	27
1.7.1 Military Data.....	27

1.7.2	Commercial/Civilian.....	28
1.8	Summary of Related Literature	30
1.8.1	Rigorous / Physical Models	31
1.8.2	Empirical Models	32
1.8.3	Registering LR DSMs to the Satellite Images	33
1.8.4	Sparse Reconstruction	34
1.8.5	Dense DSM Generation	36
1.9	Objective of the Study	36
1.10	Significance of the Study.....	37
1.11	Outline of the Thesis.....	38
2	PROJECTION MODELS.....	40
2.1	Rigorous Models.....	40
2.1.1	Kepler’s Laws.....	40
2.1.2	Orbital Parameters	41
2.1.3	The Geometrical Models	45
2.2	Empirical Models.....	48
2.2.1	Rational Polynomial Function Model (RFM)	48
2.2.2	Direct Linear Transformation (DLT)	52
2.2.3	Linear Pushbroom Camera Model (LPCM)	53
2.2.4	Affine Projection Model (APM)	56
3	PROPOSED METHODOLOGY.....	57
3.1	SRTM Registration.....	59
3.2	Sparse Correspondence Generation	60
3.3	RFM Correction.....	61
3.4	Sparse Triangulation	61
3.5	Dense DSM Generation	62
4	SRTM REGISTRATION.....	64
4.1	Available Digital Surface Models.....	64
4.1.1	SRTM.....	64

4.1.2	ASTER GDEM.....	66
4.1.3	LIDAR / LADAR	67
4.2	SRTM Registration.....	67
4.2.1	Proposed Bias Reduction Technique	71
5	SPARSE RECONSTRUCTION	77
5.1	Related Work.....	77
5.1.1	Sparse Matching	77
5.1.2	Sparse Reconstruction	85
5.2	Sparse Matching	90
5.2.1	Interest Point Detection.....	91
5.2.2	Matching.....	93
5.2.3	Elimination of Outliers	98
5.3	Sparse Reconstruction.....	100
5.3.1	RPC Bias Equalization	100
5.3.2	RPC Reconstruction	107
6	DENSE DSM GENERATION	114
6.1	Meshing.....	115
6.2	Interpolation by Integral Images.....	116
6.3	Edge-Aware Filtering.....	118
6.4	Speeded-up Turbo Pixels (SuTP)	124
7	EXPERIMENTAL RESULTS.....	126
7.1	Stereo Datasets Used in the Thesis.....	126
7.2	Ground Control Data	132
7.3	Accuracy of the Provided RPCs	133
7.4	SRTM Registration.....	135
7.4.1	Filling Gaps in the SRTM Data.....	135
7.4.2	Registration	136
7.5	Sparse Reconstruction.....	144
7.5.1	Correspondence Generation.....	144

7.5.2	Reconstruction	162
7.6	Dense DSM Generation	164
7.6.1	Performance of the Proposed Methods.....	164
7.6.2	Comparison with COTS Software	180
7.7	Execution Times.....	187
7.7.1	SRTM Registration	187
7.7.2	Sparse Reconstruction	188
7.7.3	Dense DSM Generation	192
8	CONCLUSION AND DISCUSSION	195
8.1	The Significance of the Study.....	195
8.2	Usage	198
8.3	Remaining Problems	199
8.4	Future Work	201
	REFERENCES	203
	APPENDICES	
A	EMPIRICAL MODELS FROM SRTM REGISTRATION	218
B	JACOBIAN IN THE RPC RECONSTRUCTION.....	222
	CURRICULUM VITAE.....	228

LIST OF ABBREVIATIONS

AGAST	Adaptive and Generic Accelerated Segment Test
ALOS	Advanced Land Observation Satellite
AN	Ascending Node
APM	Affine Projection Model
B/H	Base-to-Height Ratio
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
CE90	Circular Error 90
CenSurE	Center Surround Extremas
COTS	Commercial off the shelf
CS-1	CARTOSAT-1
DEM	Digital Elevation Model
DLT	Direct Linear Transformation
DN	Descending Node
DSM	Digital Surface Model
DTM	Digital Terrain Model
EOS	Earth Observation Satellite
EP	Earth Center-to-Perigee vector
ES	Earth Center-to-Satellite vector
FAST	Features from Accelerated Segment Test
FIFO	Firs-In First-Out
FOV	Field of View

GCP	Ground Control Point
GPS	Global Positioning System
GRS-80	Geodetic Reference System 1980
GSD	Ground Sample Distance
HS	Hyperspectral
ILS	Iterative Least Squares
ISPRS	International Society of Photogrammetry and Remote Sensing
KLT	Kanade-Lucas-Tomasi tracker
LADAR	Laser RADAR (a.k.a LIDAR)
LE90	Linear Error 90
LGM-TM	Template Matching on Local Gradient Maxima
LIDAR	Light Detection And Ranging
LoG	Laplacian of Gaussian
LM	Levenberg-Marquardt Algorithm
LR-DSM	Low Resolution Digital Surface Model
LPCM	Linear Pushbroom Camera Model
LWIR	Long-Wave Infrared
METI	Ministry of Economy, Trade, and Industry (Japan)
MMRG	Multimedia Research Group
MPM	Multi Primitive Matching
MS	Multispectral
MSER	Maximally Stable Extremal Regions
NASA	National Aeronautics and Space Administration
OpenCV	Open Computer Vision Library
ORB	Oriented BRIEF
P-DLT	Piecewise Direct Linear Transformation

P-PM	Piecewise Pinhole Model
RFM	Rational Function Model
RMS	Root Mean Square
RPC	Rational Polynomial Coefficients
SAR	Synthetic Aperture Radar
SIFT	Scale Independent Feature Transform
SNR	Signal-to-Noise Ratio
SPOT	Système Pour l'Observation de la Terre
SPOT HRS	SPOT High Resolution Sensor
SRTM	Shuttle Radar Topography Mission
SURF	Speeded-up Robust Features
SUSAN	Smallest Univalued Segment Assimilating Nucleus
SuTP	Speeded-up Turbo Pixels
SVD	Singular Value Decomposition
UTM	Universal Transverse Mercator
VNIR	Visual and Near Infrared
WGS-84	World Geodetic System 1984 Model
WV-1	WorldView-1

CHAPTER 1

INTRODUCTION

1.1 General Overview

Remote Sensing is the effort for extracting information on The Earth from aerial (airborne), satellite (spaceborne) or terrestrial data. The information is obtained with various imaging and signal processing techniques and used in a wide spectrum of application areas such as cartography, population studies, city planning, agriculture, land cover, mining, hydrology, defense, intelligence, meteorology, disaster monitoring, etc.

A Digital Elevation Model (DEM) is a map in raster or vector format, representing the geodetic (latitude, longitude, altitude) data of a region on The Earth. DEMs are widely used in remote sensing applications. Two types of DEMs are defined: *i*) A Digital Terrain Model (DTM) represents the earth surface without land cover or buildings (e.g. cartographic maps) and *ii*) A Digital Surface Model (DSM) represents the earth surface with land cover and buildings (terrain information is absent for tree cover and buildings). DEM generation is an active research area, especially for satellite imagery. The area of study that covers DEM generation from satellite and aerial imagery is named as *photogrammetry*.

Aerial and satellite images can be obtained with different types of sensors. Images in different wavelengths (visual and near infrared (VNIR), Multispectral (MS)) or images of different imaging techniques (LIDAR, SAR) can be used. This study is on generating DSMs from stereo satellite images of the VNIR band.

1.2 The Addressed Reader

This text addresses the following audiences:

- i) **Photogrammetry experts** who focus on DEM generation.
- ii) **Computer vision experts** who are specialized in 3D scene reconstruction
- iii) **Remote sensing experts** who use DEMs in their studies as inputs and require a deeper understanding of DEM generation.

The contents of the introduction chapter are mostly known by the photogrammetry society. This chapter addresses audiences (ii) and (iii). The remaining chapters are mostly focused on the details of the thesis study and they may be difficult to follow for the audience (iii).

The author of this text is an electrical engineer with a background in computer vision and signal processing.

In photogrammetry, the concepts and methods of computer vision are frequently used with a different terminology. However, it is observed that; the recently developed, efficient computer vision methods are not visited in the photogrammetry literature. Thus the members of the first audience group may find this text useful, since the problem is approached with the computer vision perspective.

1.3 Problem Statement

Given;

A stereo satellite image pair

$$I_1(u_1, v_1) \text{ and } I_2(u_2, v_2),$$

and the projection relations

$$u_1 = F_{11}(\phi, \lambda, h), \quad u_2 = F_{21}(\phi, \lambda, h),$$

$$v_1 = F_{12}(\phi, \lambda, h), \quad v_2 = F_{22}(\phi, \lambda, h).$$

Find the mappings

$$\Phi(u_1, v_1), \Lambda(u_1, v_1) \text{ and } H(u_1, v_1)$$

where,

u_1, v_1 : Row and column indexes for image 1

u_2, v_2 : Row and column indexes for image 2

ϕ : Geodetic latitude of the observed point

λ : Geodetic longitude of the observed point

h : Geodetic altitude of the observed point

Φ : Reconstructed latitude of the observed point

Λ : Reconstructed longitude of the observed point

H : Reconstructed altitude of the observed point

As seen in the above problem statement, the objective is to assign latitude, longitude and height values for the reference image only. Obviously, the output can be generated on a latitude-longitude grid with additional post-processing steps.

Note that, apart from the computer vision convention, here we use u for the vertical axis and v for the horizontal axis in the image domain. Often, X , Y and Z are used for the **normalized** latitude, longitude and height,

respectively. Normalized image domain coordinates are also used frequently in the projection relations.

Although the projection relations are shown here as explicit functions for simplicity, they may not be explicit in reality. In some rigorous projection models, the projections are implemented as procedures rather than analytical functions (*Gupta 1995, Jacobsen 1997*).

1.4 Error Measures

In photogrammetric DEM generation, the following reconstruction errors are defined for a specific point in the image:

$$\begin{aligned} e_{\phi}(u_1, v_1) &= \Phi'(u_1, v_1) - \Phi(u_1, v_1), \\ e_{\lambda}(u_1, v_1) &= \Lambda'(u_1, v_1) - \Lambda(u_1, v_1), \\ e_h(u_1, v_1) &= H'(u_1, v_1) - H(u_1, v_1) \end{aligned} \tag{1.1}$$

where,

- $\Phi'(u_1, v_1)$: actual geodetic latitude (ground truth) for pixel (u_1, v_1) ,
- $\Lambda'(u_1, v_1)$: actual geodetic longitude (ground truth) for pixel (u_1, v_1) ,
- $H'(u_1, v_1)$: actual geodetic altitude (ground truth) for pixel (u_1, v_1) ,
- $\Phi(u_1, v_1)$: reconstructed geodetic latitude for pixel (u_1, v_1) ,
- $\Lambda(u_1, v_1)$: reconstructed geodetic longitude for pixel (u_1, v_1) ,
- $H(u_1, v_1)$: reconstructed geodetic altitude for pixel (u_1, v_1) .

The errors are measured at Ground Control Points (GCPs) which are obtained from field measurements. If the generated DEM is defined on a latitude-longitude grid, airborne LIDAR data can also be used.

The error expressions in the geodetic coordinate system (see Section 1.5.5) are not ergonomic, since the errors are mostly in microdegrees and the

same longitude error (in degrees) corresponds to different displacement errors (in meters) on different latitudes. The errors are preferred to be in meters.

For DEMs, two separate error definitions are of importance: the *planimetric error* (geolocation error) and the *elevation error* (height error). The *planimetric error* is defined as the error in the XY-plane, which is tangential to the reference ellipsoid at the actual latitude and longitude (i.e., the distance between the actual and the estimated transversal positions). The *elevation error* is defined as the difference between the actual and the estimated altitudes of that point (Figure 1.1).

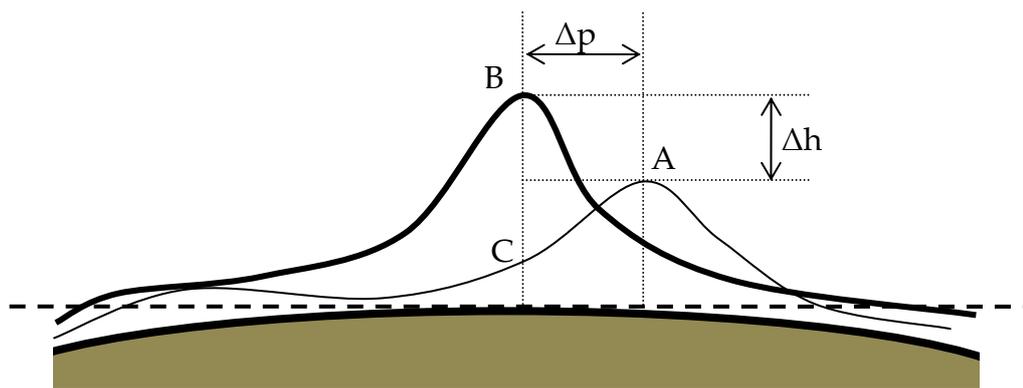


Figure 1.1. Plannimetric error (Δp) and elevation error (Δh) for the simplified (2-D) imaging geometry. The actual world point B is reconstructed as if it were at A. The hypothetical tangential plane is denoted by the dashed line.

Note that the computations of plannimetric and elevation errors are ambiguous, since any DEM states a height value for every location. For instance; in Figure 1.1, for the latitude and longitude of point B, the DEM has a value (the height of point C). In a more radical case, if the DEM is flat everywhere; the plannimetric error cannot be measured.

The ambiguity is eliminated if the points of interest are on a satellite image. A DEM that is originated from *photogrammetry* states the geodetic coordinates (lat, lon, h) for the observed image pixels whose latitudes and longitudes can be determined accurately in the field. Thus, the errors are more sound, when they are given with a mapping from the reference image to the object domain.

In the photogrammetry literature, the most frequently reported errors are the following:

- i) Mean absolute planimetric error
- ii) Mean elevation error
- iii) Standard deviation of the planimetric and elevation errors,
- iv) RMSE for planimetric and elevation errors
- v) Minimum/Maximum elevation and planimetric errors
- vi) Circular planimetric error; CE90
- vii) Linear elevation error LE90

The last two error definitions are originated from the cumulative distribution function of the error magnitudes. LE90 states the error value, at which the cumulative distribution function reaches 90%. The same is true for CE90. The elevation error is 1D while the planimetric error is 2D. It is sufficient to take the absolute value for the elevation error, but the magnitude of the planimetric error is defined using the L_2 norm, which corresponds to the radius of the smallest circle that encloses the error vector.

1.5 Satellite Imagery

Satellite images (also called spaceborne images) are the images taken by Earth Observation Satellites (EOS). There are various types of EOSs, with

different properties and sensor resolutions depending on the applications they are used for and the technology they are built with.

All EOS orbits lie in planes that are close to the earth's axis of rotation. That is, all EOSs pass over the North and the South Poles. Since The Earth rotates from the West to the East, the satellites fly over different regions of The Earth at each tour in the orbit.

Before proceeding further, it is necessary to introduce the basics and terminology of the EOSs.

1.5.1 The Imaging Scenario

All EOSs are equipped with a specific sensor structure named as "*pushbroom sensor*". A pushbroom sensor is a 1-D array of sensors combined with proper optical structures for zooming. The image acquisition strategy is demonstrated in Figure 1.2 (the scenario is over-simplified and the scales are exaggerated for the sake of clarity). The satellite images are obtained by time-sampling the sensor output while the satellite moves in its orbit. Due to this image acquisition strategy, pushbroom sensors are also named as "*digital scanners*" (Toutin 2001).

Figure 1.2 can be explained as follows: The sensor is used as a single-row camera which takes successive snapshots while the satellite moves in its orbit. The integration time and the look direction (*attitude*) of this single row camera are adjusted such that, when all of the snapshots are combined in the order of acquisition time, a complete image of the observed region is obtained. To avoid gaps on the ground, a slight overlapping is used between adjacent rows (Figure 1.2a), which is ignored in many studies.

Two directions are defined in the pushbroom scanners:

i) The *along-track* direction: The direction of flight in the orbit

ii) **The cross-track direction:** The direction of the sensor array which is perpendicular to the along-track direction. In some references, this direction is also referred as the *across-track* direction.

The spatial resolution of the EOSs is stated using the Ground Sample Distance (GSD). The distance that can be observed in the cross-track direction (span of the sensor array on the earth surface) is defined as the *swath* width.

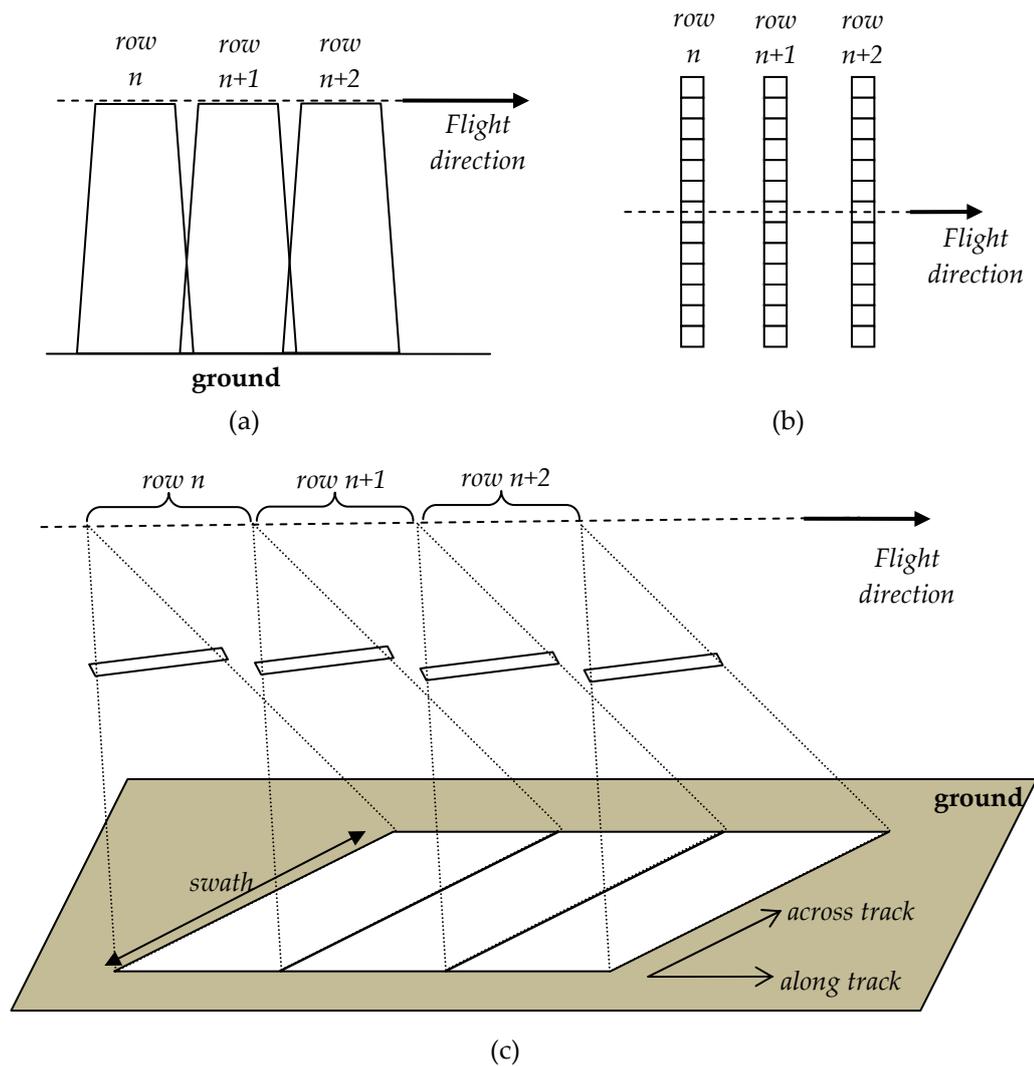


Figure 1.2. Pushbroom imaging scenario: Side view (a), Top view (b), and oblique view (c). Each line (row) of the image is acquired at a different satellite position. Curved brackets represent the positions during the integration time for each row.

The camera may be static or rotating (agile). In the rotating camera configuration, it is possible to take samples against the flight direction in the *asynchronous mode* (Figure 1.3). This strategy is used to obtain higher resolution images with some sacrifice for the scanning area per unit time (Poli 2012). Modern rotating cameras are also capable of scanning in the across-track direction, resulting in a scan in the east-west direction (Jacobsen 2003).

The imaging convention for the satellite images puts each scanning line as a row in the image. Thus, satellite images reflect perspective projection along the x-axis and pushbroom projection along the y-axis.

In the remote sensing literature, for the image domain coordinates, the terms *line* and *sample* are frequently used for vertical (u) and horizontal (v) coordinates, respectively. *Line* and *sample* always take integer values.

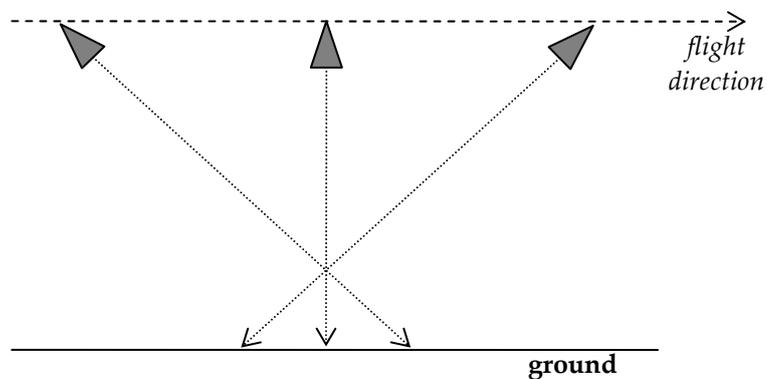


Figure 1.3. Imaging against the flight direction with a rotating camera.

Stereo Imaging

Stereo imaging is performed by acquiring the image pair at two separate satellite locations. Two types of stereo imaging strategy can be used: *i) Along-track* (or in-track) stereo and *ii) across-track* stereo. In the *Along-track* stereo case, image pair is acquired in a single pass above the ground location (Figure 1.4). In the *across-track* stereo case, the image pair is acquired in two separate passes (Figure 1.5).

Along-track stereo images can be taken with two different approaches:

- i) Multiple Static Cameras** (Figure 1.4a): The cameras are fixed to the satellite body with carefully determined look directions so that the satellite obtains the stereo images of a scene at some fixed positions in the orbit.
- ii) Single Rotating Camera** (Figure 1.4b): The satellite is equipped with a rotating camera and the electro-mechanic gear for directing the camera to a desired location. Thus, the camera can be aimed freely by adjusting its roll and pitch (the roll and pitch are defined for the satellite, in accordance with the flight dynamics convention). The stereo imaging with the single rotating camera is performed by turning the camera to the same spot at different positions in the orbit.

The configuration with multiple static cameras has the advantage that; compared to rotating camera systems, more image pairs can be collected per unit time. Additionally, it is not required to measure the camera orientation with respect to the satellite. On the other hand, the single rotating camera configuration allows the user to obtain images with different base-to-height (B/H) ratios (see Section 1.5.4). This configuration also enables both along-

track and across-track stereo, while static cameras are designed to perform either across-track or along-track stereo.

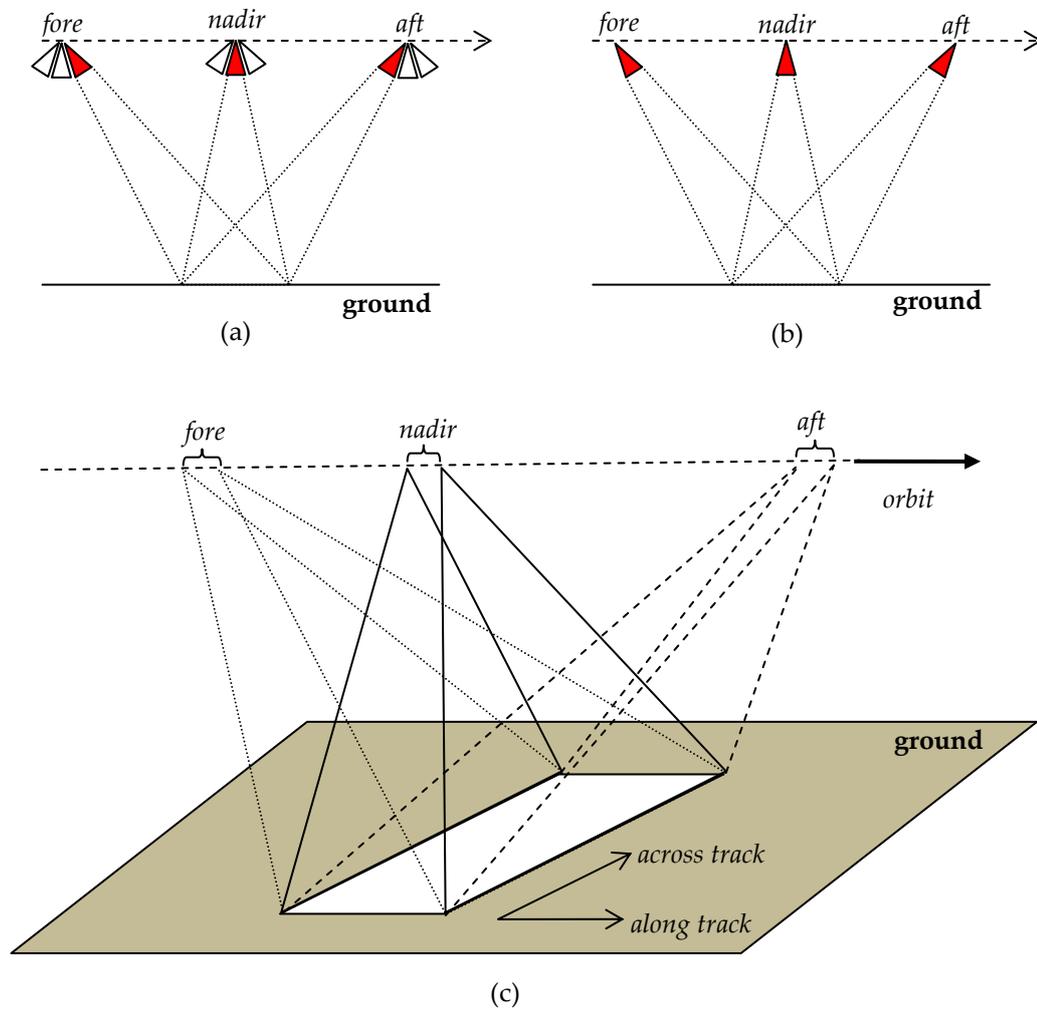


Figure 1.4. Along-track stereo; (a) multiple static cameras, (b) rotating camera and (c) oblique view

Across-track stereo images are obtained using adjacent orbits by aligning the camera to the across-track direction in the tour after the nadir imaging tour. Images taken at different dates can also be used as across-track stereo, provided that the satellite positions of the two images correspond to

sufficient B/H ratio. The minimum time difference between across-track stereo images is one period in the orbit, which is 90 to 100 minutes for EOSs.

Along-track stereo is preferred to across-track stereo since the effects of illumination and/or scene change is minimal.

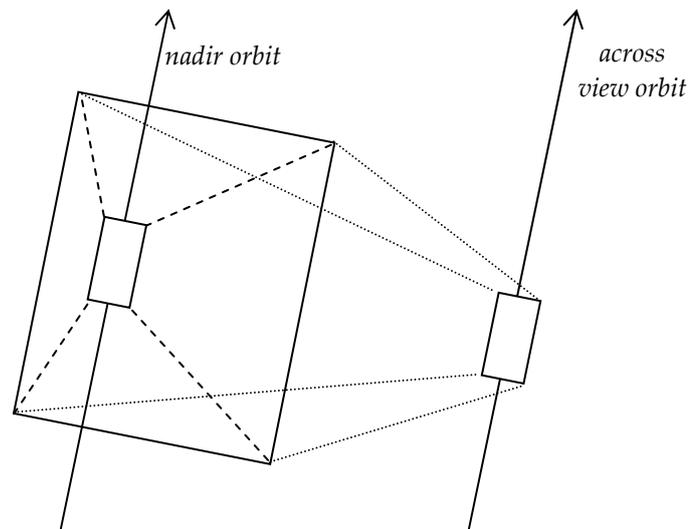


Figure 1.5. Across-track stereo imaging. The EOS turns his camera in the across-track direction while it moves in an off-nadir orbit

1.5.2 The Collinearity Equations

The collinearity equation is the fundamental equation for all projective camera systems. Given an imaging system with a focal length f , the 3-D point (*object point*) at (X, Y, Z) that is observed at pixel (u, v) lies on a line (*projection line*) which is defined by the optical center (of the imaging system) and the sensor (pixel) that acquires the pixel data at (u, v) . The equations describing this fact in mathematical form (*collinearity equations*) are (Toutin 2004):

$$\begin{aligned}
u &= (-f) \frac{m_{11}(X - X_0) + m_{12}(Y - Y_0) + m_{13}(Z - Z_0)}{m_{31}(X - X_0) + m_{32}(Y - Y_0) + m_{33}(Z - Z_0)} \\
v &= (-f) \frac{m_{21}(X - X_0) + m_{22}(Y - Y_0) + m_{23}(Z - Z_0)}{m_{31}(X - X_0) + m_{32}(Y - Y_0) + m_{33}(Z - Z_0)}
\end{aligned}
\tag{1.2}$$

where f is the focal length of the imaging system, (u, v) are the image plane coordinates, (X_0, Y_0, Z_0) are the projection center coordinates and (X, Y, Z) are the *object point's* coordinates. The nine parameters $m_{11}..m_{33}$ are the parameters of the orthogonal rotation matrix between the image coordinate system (that defines (u, v, k) , 3rd dimension obeying the right-handed system) and the world coordinate system (that defines (X, Y, Z) and (X_0, Y_0, Z_0)).

In the traditional perspective projection (pinhole) model, the optical center coordinate and the image plane is fixed for all pixels (rows). Therefore, the rotation matrix parameters are the same for all points in the image plane, and a single projection matrix can be used. On the other hand, for the pushbroom satellite sensors, the projection center and the image plane are different for each scan line. Consequently, the nine parameters are different for different image lines.

Collinearity equations give the image coordinates of an *object point* with known world coordinates, given the sensor position and orientation (image plane normal vector for the corresponding image row). However, they are not invertible. In other words, given the camera position and orientation, the world coordinates of the imaged point cannot be obtained from a single image (2 equations with 3 unknowns: X, Y, Z) since the object point can be anywhere on the *projection line*.

1.5.3 The Coplanarity Constraints

Similar to the multiview case for the pinhole cameras, to obtain the 3D world coordinates of an *object point*, at least two images are required. Assuming a stationary scene, the *object point* can be localized by intersecting the corresponding *projection lines* of the two images. The constraint for determining the 3D point is called the *coplanarity* condition (equivalent to the *epipolar constraint* in the computer vision terminology) and obtaining the 3D world coordinate is called *triangulation*. The name “coplanarity” comes from the fact that the object point and two optic centers define a plane. When equation (1.2) is converted to the form

$$\begin{aligned} \frac{X - X_1}{a} &= \frac{Y - Y_1}{b} = \frac{Z - Z_1}{c} \\ \frac{X - X_1}{d} &= \frac{Y - Y_1}{e} = \frac{Z - Z_1}{f} \end{aligned} \quad (1.3)$$

then, the intersection of the two lines are given by

$$\begin{aligned} X &= \frac{ad(Y_2 - Y_1) + bdX_1 - aeX_2}{bd - ae} \\ Y &= \frac{b(X - X_1)}{a} + Y_1 \\ Z &= \frac{c(X - X_1)}{a} + Z_1 \end{aligned} \quad (1.4)$$

Here, (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) are the camera centers for the two images. Note that, for the satellite imaging case, each image row is acquired with a different optic center and therefore, the term “optic center” refers to the optic center for the row that the world point is observed.

This simple relation does not reflect the real life situation, since the projection lines do not intersect in space due to various sources of noise (spatial sampling, projection model imperfections, errors in matching, limited precision, etc).

1.5.4 Satellite Imaging Systems

EOSs are required to provide quality images with a sub-meter GSD over a wide swath while they are moving in an orbit 400-700 km above the ground. The required accuracy levels for the DEMs and the orthophotos are at most 1 GSD. Thus, EOS design and data interpretation are very complicated tasks.

The following remarks are useful for summarizing the satellite imaging:

i) Design Challenges: For satellite imaging systems, the swath is desired to be large, and the spatial resolution is desired to be high. Providing the two properties at the same time is a difficult task. A camera providing both properties needs to have a large FOV. The size of the satellite (the aperture size) is limited; therefore the focal length will be small for a large FOV camera. This will lead to limited optical zooming capacities. At this point, high resolution can only be obtained by smaller sensor cells, but *this* will lead to lower SNR. The effects of this dilemma can be seen in Table 1.2

In the practical systems, smaller FOVs are used to meet the resolution needs. Thus, in general, as the Ground Sample Distance (GSD) decreases (improves), the swath also decreases (Table 1.2).

The need for high spatial resolution (~1m) in the *along-track* direction and the satellite linear velocity bounds the sampling interval (*dwell time*) (e.g. SPOT: 1.504ms) (Gupta 1995). On the other

hand, due to the high altitude of the satellites, the optical systems are required to have significant zooming capacities (high focal length) (*Piwowar 2006*). As the focal length increases, the lens attenuation of input light increases, and therefore, in order to have a short exposure time, special optical and sensor design (larger aperture, larger sensors, etc) is required. For instance, IKONOS satellite sensor size is; pan: 12 μm and color: 48 μm (*Piwowar 2009*) while a Canon 1D Mark III sensor size is 7.2 μm (*Canon 2008*)

ii) Full coverage: Because of Earth's rotation, the satellite passes over all points on the Earth. This allows a satellite to observe the entire Earth surface. On the other hand, an EOS cannot cover the entire Earth surface in a single day. An EOS completes at most 20 orbits per day with a limited swath (10-200km).

iii) Extremely low temporal resolution: It is not possible for an EOS to observe a specific location on Earth continuously. This is possible only if the satellite is on the equatorial plane, at a specific altitude and directed to the desired location (geostationary satellites, (*Clarke 1945, Kelso 1998*)). However the orbital radius for geostationary satellites is 35786 km (*Kelso 1998*), which is not feasible for high resolution Earth observation (see Table 3 for EOS orbit radii). The EOS re-visit period for a specific location on the earth surface is in the order of days. Due to the 1-D structure of the sensor; at each pass, only a single image or a few images (in-track stereo) can be obtained for a specific location. Therefore, a single EOS cannot be used to monitor fast temporal changes in a specific location.

iv) Base-to-Height (B/H) Ratio: Since the satellite altitude is high, for accurate DEM generation, the displacement (base distance) between

the (stereo) acquisition positions must be comparable to the satellite altitude (height). A B/H of 1.0 is accepted to be “good” (*Toutin 2001*) and this value is taken as a goal in camera design.

- v) **Complicated models:** The imaging scenario includes orbital motion and sensor orientation parameters, the earth’s ellipsoid model and the earth’s rotation. Thus, rigorous modeling of the imaging scenario is a complicated task (the computations in Section 2.1.2 must be performed for each image line).
- vi) **Dependence on ephemeris accuracy:** The altitudes of the EOSs (a few hundred kilometers) require very precise measurements of sensor position and orientation (*attitude*). For example, a 2-arc-second error in IKONOS sensor orientation results in at least 6.6 meters of displacement error on Earth’s surface (*Grodecki 2003*). For this reason, satellites are equipped with special hardware (star trackers, gyroscopes, encoders, GPS) for measuring the small changes in the satellite position and attitude. The readings (*ephemeris*) obtained from the measurement equipments are sent to the ground station for each image.
- vii) **Computational complexity:** The number of sensors (~6K-30K) in the 1-D array is larger than the number of sensors (~1K-4K) in a row of a traditional 2-D digital sensor. This strategy provides very large (150 to 900 megapixel) images. The operations to be performed on these images involve subsequent expensive processing steps, leading to high computational and memory complexity.
- viii) **High data costs:** Satellite imaging is an expensive task. Even the least processed data are costly. The difficulties in sensor modeling and the dependence on the physical sensor parameters bring further

expense to the processed spaceborne products (*Fraser 2001, Fraser 2002*).

1.5.5 Coordinate Systems

The satellite position and ground points can be represented by different coordinate systems. For satellite dynamics computations, an Earth-centered, orthogonal coordinate system is used (*geocentric* coordinates). In the geocentric coordinate system, z-axis points towards the North Pole, x-axis points towards the 0° longitude, and y-axis is decided according to the right-hand rule (points towards 90° East). Precise reference frames are defined by various versions of the International Terrestrial Reference Frame (*Arias 1994, Boucher 1999, Altamimi 2002, Altamimi 2007, Altamimi 2011*).

On the other hand, in cartography, *geodetic* coordinate system (latitude, longitude, height) is used (GRS-80 (*Moritz 1992*), WGS-84 (*U.S. Department of Defence 1997*)). Geodetic coordinate systems use reference ellipsoid models for the surface of The Earth at sea level (see section 1.5.6), and define the height of any geographical point as the distance from this point to the model ellipsoid's surface, in the surface normal direction (Figure 1.6).

The point on the Earth, whose surface normal points to the satellite is called the *nadir point*. In other words, the nadir point is the Earth point that is closest to the satellite. The satellite's footprint is defined using the nadir points.

Note that, as seen in Figure 1.6, the surface normal does not point towards The Earth's center (except on The Equator and the poles). Therefore, the conversion from geodetic coordinates to geocentric Cartesian coordinates is different from the traditional spherical-to-Cartesian conversion. For the same reason, geocentric and geodetic definitions of latitude are different.

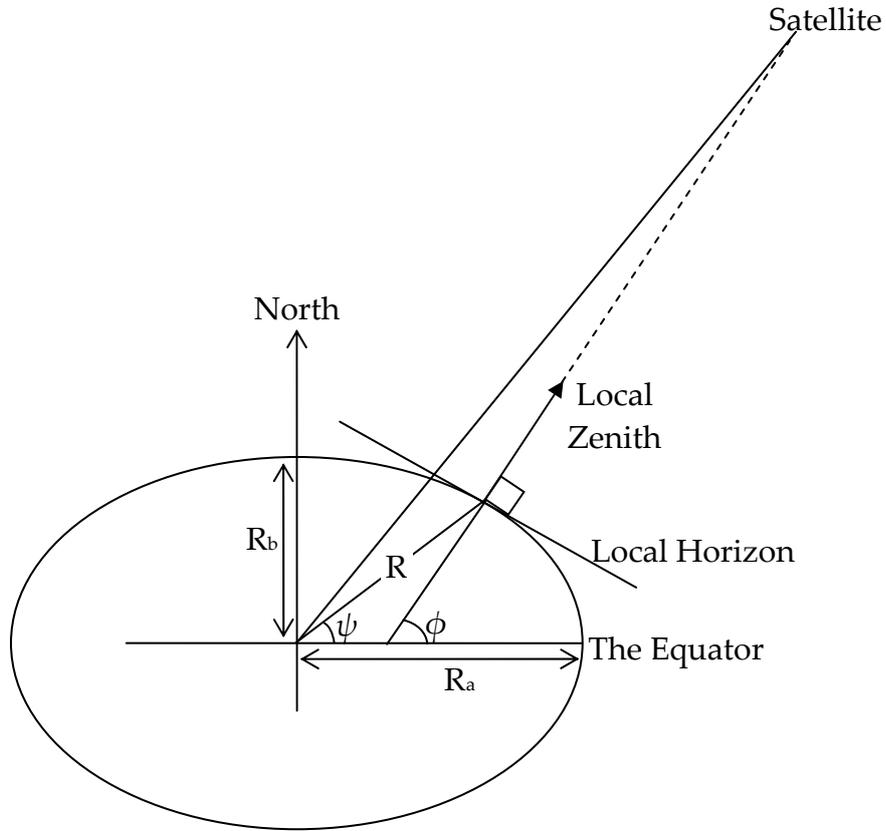


Figure 1.6. Geodetic and geocentric coordinates of a ground point (a), and geocentric satellite coordinates (b) (Kelso 1996)

The transformations from geocentric latitude (ψ) and longitude (η) to the geodetic counterparts (ϕ, λ) are given by (Gupta 1995):

$$\begin{aligned} \lambda &= \eta \\ \phi &= \tan^{-1} \left(\frac{R_a^2}{R_b^2} \tan(\psi) \right) \end{aligned} \quad (1.5)$$

where R_a is the semi-major axis and R_b is the semi-minor axis of the reference ellipsoid. Obviously, the geocentric pair (ψ, η) represent the spherical

coordinates (azimuth and polar angles), with the only difference that, for the polar angle, the equatorial plane is taken as 0° . Note that these two coordinate systems are Earth-fixed (that is, they rotate with The Earth).

The geodetic (ϕ, λ, h) and geocentric (ψ, η, R) coordinates can be converted to Earth-centered, Earth-fixed Cartesian Coordinates (X, Y, Z) with (Gupta 1995)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} R \cos(\psi) \cos(\eta) \\ R \cos(\psi) \sin(\eta) \\ R \sin(\psi) \end{bmatrix} \quad (1.6)$$

and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (N \cos \psi + h \cos \phi) \cos \lambda \\ (N \cos \psi + h \cos \phi) \sin \lambda \\ N \sin \psi + h \sin \phi \end{bmatrix} \quad (1.7)$$

Here, N is

$$N = \frac{R_a R_b}{R_b^2 \cos^2 \psi + R_a^2 \sin^2 \psi} \quad (1.8)$$

Note that, the transformation from geodetic coordinates to geocentric Cartesian coordinates by (1.6) and (1.7) requires the geocentric coordinates, which can easily be obtained by using (inverting) the relations in (1.5)

However, satellites do not rotate with The Earth. Therefore, registering satellite position with the geodetic coordinates is required. The details are presented in Section 2.1.2 (orbital parameters).

Earth-centered Cartesian coordinate system is rarely used for describing the imaging geometry since it requires extra precision for numerical stability. Still, in rigorous models, usage geocentric coordinates is inevitable. On the other hand, the Cartesian coordinate system can be used locally. The *Universal Transverse Mercator* (UTM) is a 2-D Cartesian coordinate system, in which the earth surface is divided into 14 regions which are assumed to be planar (*Hager 1989*). In each region; the coordinates are defined as 2-D Cartesian coordinates relative to the origin defined for the UTM region. Although UTM defines the transverse plane coordinates only, the height information may still be represented in this coordinate system relative to the UTM plane. UTM is generally used as a projection plane and is rarely used in photogrammetry as the object domain coordinate system.

1.5.6 Earth Models

Numerical description for the satellite imaging scenario through the projection models requires a reference Earth model. The Earth is modeled as an ellipsoid, which is defined by two parameters: The *semi-major axis* (R_a) and the *semi-minor axis* (R_b). Every Earth model defines a reference surface, in reference to which the coordinates of a point on the ground is described in the geodetic or the geocentric coordinate system (see Figure 1.6).

Several ellipsoids are defined throughout the years. The parameter values for the different reference ellipsoids are given in Table 1.1. The most widely used reference ellipsoid model is the WGS84 model (*NIMA 2000*).

Table 1.1 Reference ellipsoid parameters for different ellipsoidal models

Name	R_a (km)	R_b (km)
Clarke 1866*	6378.2064	6356.5838
Bessel 1841*	6377.397155	6356.078965
International 1924*	6378.388	6356.9119
GRS 1980*	6378.137	6356.7523141
WGS 1984 (NIMA 2000)	6378.137	6356.7523142

*The Defence Mapping Agency, 1983

1.6 Data Types

Various types of data are used in remote sensing for different purposes. This section presents a general overview of the available data and their usage, and marks the data types that are used in photogrammetric DEM generation studies.

1.6.1 Monochrome / Panchromatic Images

Panchromatic (pan) images are the most common type of data provided by satellite image providers. The sensors used to obtain the pan images have a wider spectral response than that of the multispectral sensors (Figure 1.7 and Figure 1.8). Resolution is generally 4 times higher than colored (generally VNIR) images (QUICKBIRD pan: 0.6m, colored: 2.4m (*DigitalGlobe 2008*), IKONOS pan: 1m, colored: 4m (*Pitovovar 2006*), ALOS pan (PRISM): 2.5m, colored (AVNIR): 10m (*Osawa 2004*)). Pan images are the most widely used data in DEM extraction.

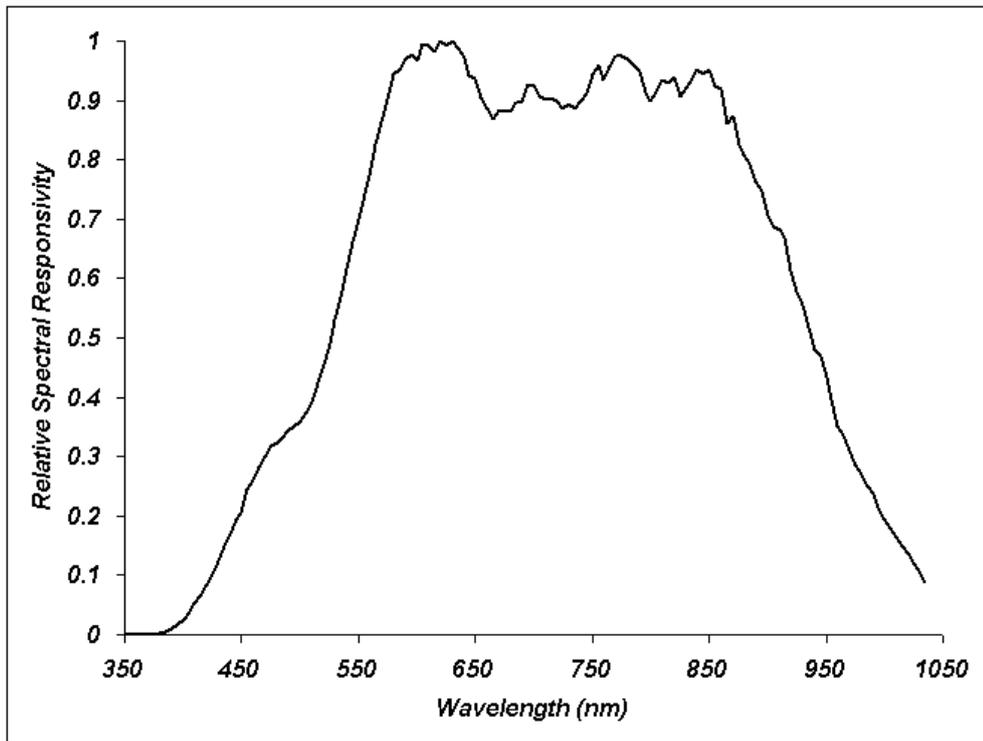


Figure 1.7. Spectral response for IKONOS pan sensor. Taken from *GeoEye Web Page*

1.6.2 Multispectral / Colored Images

Multispectral (MS) images are multi-channel images, each channel containing sensed information in a specific spectral range. A typical MS image usually covers the visible and near-infrared (VNIR) bands. MS data may have additional bands (e.g. 7 bands in LANDSAT ETM+). Resolution is lower than the pan images. Image providers may sell pan-sharpened products, which are generated by increasing the colored image resolution using the information in the pan images. Although the usage of MS images is not very common (due to the superior pan image resolution), they are useful in image segmentation, which can be used in dense DEM generation.

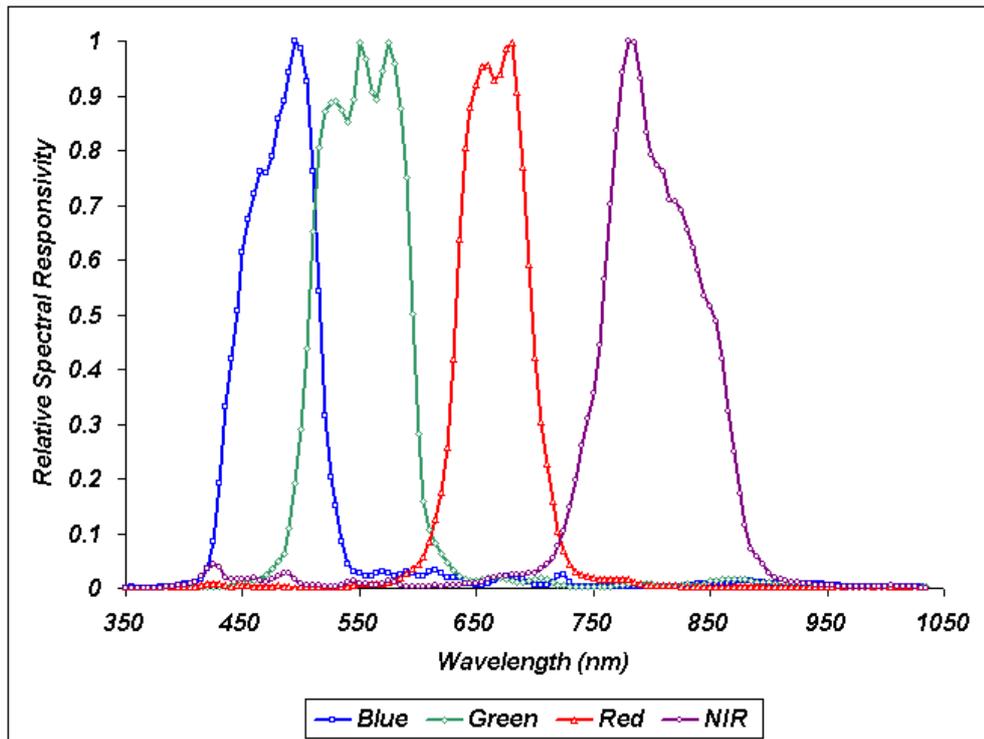


Figure 1.8. Spectral response for IKONOS MS sensors. Taken from *GeoEye Web Page*

1.6.3 Hyperspectral Images

Hyperspectral (HS) data include the images for a large number of frequency bands. For instance, 220 bands are provided by Hyperion, with 30m spatial resolution (*USGS EO-1 website*). Although hyperspectral data have lower spatial resolution than pan or MS images of modern satellites, they provide powerful data for various purposes (agriculture, disaster monitoring, etc.). The details of HS imaging are beyond the scope of this study, since HS data are not used for DSM generation due to its limited spatial resolution.

1.6.4 Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) is a special kind of radar, which aims at obtaining a reflectance image for a specific radar wavelength, by observing the echoes of emitted radar signals. The required large receiver antenna is obtained virtually by emitting a pulse train as the receiver moves, and combining the data gathered at different positions. Thus, a “synthetic” aperture is generated. SAR can provide sub-meter resolution, but the images have significant speckle noise, which causes difficulties in processing. Spaceborne SAR is also practical (e.g. ALOS PALSAR, TerraSAR-X, RADARSAT-2) down to 1m resolution (*German Aerospace Center*). One major advantage of SAR imaging is the ability to acquire images at night and seeing even through a thick cloud cover.

For DSM generation, a specific form of SAR, namely the Interferometric SAR (InSAR/IfSAR) is used. InSAR can provide accuracy figures down to centimeters (*Helz 2005*).

Spaceborne stereo SAR can be used for DSM generation with similar techniques for correspondence generation (*Toutin 2012*). However; the imaging geometry, the properties of the SAR images and the camera models for SAR are different from those of the electro-optic counterparts, and therefore, SAR is beyond the scope of this study. (see *Toutin 2000* for an introduction on SAR imaging).

1.6.5 Ground Control Points

A Ground Control Point is a point on the earth surface whose geodetic coordinates are measured very accurately and whose location in the satellite

image is also known. For the stereo case, the location of the GCP must be known in both images for being useful in stereo processing.

GCPs are crucial in many photogrammetry studies that demand accuracy. The most frequent usage is in the bundle adjustment, where the GCPs are used as anchors to correct the errors in the projection models (Grodecki 2003).

GCPs are also used in accuracy assessment for the projection models and generated DEMs. For proper error analysis, the GCPs are generally divided into training and test groups and the errors are measured only at the test points.

LIDAR (often used as LADAR in the military context) aims to obtain the elevation directly from time of flight of the sent laser; therefore LIDAR imaging is a different way of DEM generation, and named as *laser altimetry*. Airborne LIDAR is frequently used as a reference for accuracy measurement in the recent studies on DSM generation from satellite images (Tsutsui 2007).

1.6.6 Low Resolution Digital Surface Models

Low Resolution Digital Surface Models (LR DSMs) are DSMs that were obtained by specific missions or extracted from previously recorded data. The most popular datasets are SRTM and ASTER GDEM; which are freely available, raster format datasets.

SRTM

The Shuttle Radar Topography Mission (SRTM) is a 11-day mission performed by NASA in February 2000, to obtain a digital topographic database of the Earth by a radar system that was flown by the U.S. Space

Shuttle Endeavour. The GSD of SRTM is 30m in the U.S.A. and 90m elsewhere. The SRTM data is distributed free of charge from the USGS server in $1^\circ \times 1^\circ$ patches on a latitude-longitude grid for the land between the latitudes 56° S and 60° N. Current version of the SRTM data is SRTM3. For more details on SRTM, see Chapter 3.

ASTER GDEM

ASTER Global Digital Elevation Model (ASTER GDEM) was developed jointly by the Ministry of Economy, Trade, and Industry (METI) of Japan and NASA. The DEM data are extracted from the large image database of ASTER images. Its properties are summarized as follows:

The ASTER GDEM covers land surfaces between 83° N and 83° S and is composed of 22,600 1° -by- 1° tiles. Tiles that contain at least 0.01% land area are included. The ASTER GDEM is in GeoTIFF format with geographic lat/long coordinates and a 1 arc-second (30 m) grid of elevation postings. It is referenced to the WGS84/EGM96 geoid. Pre-production estimated accuracies for this global product were 20 meters at 95 % confidence for vertical data and 30 meters at 95 % confidence for horizontal data (*J Space Systems 2011*).

1.7 Sources of Satellite Data

1.7.1 Military Data

The properties of the military satellites and military spaceborne data are classified. Thus, the information on military data is very limited.

USA is the leading country in military spaceborne earth observation for the last five decades, starting with the Corona satellites in 1959 (*Perry 1973*). USA remains to have the largest fleet (122 satellites) for earth observation (*USGS 2012*). The USA military satellites are believed to resemble the

Hubble Space Telescope; which has a 2.4m aperture, 57.6m focal length and 0.043 arcseconds resolution in the visible band (*Simpson 2010, Space Telescope Science Institute 2011*). Such imaging ability corresponds to approximately 14cm GSD, if the camera is directed towards The Earth.

Russia and China follow USA in military spaceborne reconnaissance (*Space Telescope Science Institute 2011*). Surprisingly, even the newly launched Russian satellites carry film cameras whose recordings are sent to the ground using canisters that are dropped from the satellites.

European military satellites are also increasing in number. One of the themes in The European Commission FP7 Framework is “Space”, with a 1.4 billion € budget. Several projects with military aspects are conducted within this framework (LIMES, G-MOSAIC, NEWA, SSA, EU DEM). (*EU Joint Task Force 2010*)

1.7.2 Commercial/Civilian

Civilian satellites provide different types of images with varying prices, depending on the processing level of the satellite. The raw images (given different names by different image providers, here we call *level 0*) contain the unprocessed information received from the satellite. This class of images include the effect of radiometric (non-identical sensor, atmospheric distortions, etc) and lens distortions, as well as the geometric distortions (earth curvature, satellite orbit, attitude drift, etc). *Level 1* images are corrected (calibrated) images for radiometric distortions, without any processing for geometric distortions. Higher level images (geo-corrected, geo-referenced, ortho-rectified) can also be obtained from the image providers, as well as DSMs obtained from these images. Level 1 data are the most common inputs for photogrammetry studies.

Information on the orbital parameters and imaging properties of a subset of civilian EOSs are given in Table 1.2. More detailed information can be found in *Jacobsen 2002, Dowman 2003, Gleyzes 2003, Baudoin 2004, Osawa 2004, Piwovar 2006, DigitalGlobe 2008, Ghadyani 2008, Toutin 2009, Poli 2012, Satellite Imaging Corp. Web Site*

Table 1.2. Properties of differen EOSs

SATELLITE	ALOS PRISM/AVNIR2	IKONOS 2	QUICKBIRD 2	SPOT 5
Resolution (m)	2.5 Pan 10 VNIR	0.8 Pan 4 VNIR 1.0 PS*	0.6 Pan, 2.4 VNIR	2.5 Pan 10 VNIR 20 SWIR
Swath (km)	70 nadir 35 fore/aft	11.3	16.5	60
Image Width (pixels)	28000 pan 14000 MS	13816 pan 3454 VNIR	27000 pan 6700 VNIR	12000 pan 6000 MS
Cameras	+24°, 0°, -24°	Up to +/- 45°	Up to +/- 30°	+20°, -20°
Sensor	Pan, VNIR 4 band	Pan, VNIR 4band	Pan, VNIR 4 band	Pan, MS 4 band
Altitude (km)	691.65	680	470	822
B/H	1.0	variable	variable	0.8
Quantization (bits/pixel)	8	11	11	8
Orbital inclination (°)	98.16	98.10	98	98.7
Orbital Period (min)	98.74	98.33	93.4	101.4
Recurrent period (days)**	46 (2)	144 (3-5)	1-3.5	26 (2-3)
Launch	2006	1999	2001	2002
Country	Japan	USA	USA	France

* PS: pan sharpened product obtained by increasing the resolution of the MS image using panchromatic data.

** The term in the parenthesis is the sub-cycle period when the satellite passes through a point very close to the recurrence location

Table 1.2 (cont'd). Properties of different EOSs

SATELLITE	GeoEye-1	ASTER	EROS B	LANDSAT 7 ETM+	IRS-P5 Cartosat
Resolution (m)	0.41 Pan 1.65 VNIR	15 VNIR 30 SWIR 90 LWIR	0.82 Pan, 4 VNIR	15 pan 30 VNIR	2.5 Pan
Swath (km)	15.2	60	16.4	185	27 / 30
Image Width (pixels)	37544 Pan 9386 VNIR	4000 VNIR 2000 SWIR 670 LWIR	20000 Pan 5000 MS	12300 Pan 6150 VNIR 6150 SWIR 3050 LWIR	12 288 Pan
Cameras	Up to +/-45°	0° +27.2°	Up to +/-45°	0°	+20°, -5°
Sensor	Pan VNIR 4 band	MS 14 band	Pan, MS 4 band	Pan, MS 16 band	Pan
Altitude (km)	681	705	600	705	617
B/H	variable	0.6	variable	N/A	0.62
Quantization (bits/pixel)	11	8 VNIR 8 SWIR 12 LWIR	10	8	10
Orbital Inclination (°)	98	98.3	97.4	98.2	98.9
Orbital Period (min)	98	99	94-96	99	97.1
Recurrent period (days)	3	16	1-3	16	5
Launch	2008	1999	2004	1999	2005
Country	USA	USA/Japan	Cyprus / Israel	USA	India

1.8 Summary of Related Literature

The DEM generation process includes a number of steps (which are summarized in Section 1.11), each having a different subset of literature related to various approaches. Thus, for the sake of clarity, the related

literature is presented at the beginning of each chapter. Here, we provide a brief introduction with an organization that matches with the outline of the thesis work.

The usage of photogrammetry in Digital Elevation Model generation has a broad literature, starting from the early 70's (for analog cameras). With the improvements in the spatial resolution thanks to the digital satellite images, the errors that were previously neglected have gained importance and are dealt with more rigor (*Toutin 2004*).

The most critical step in DEM generation is obtaining an accurate projection model. The accuracy of the reconstructed DEM is sensitive to errors in the projective model parameters. Thus, the photogrammetry literature is rich for camera calibration and accuracy assessment studies. Since each EOS has a unique design and imaging scenario, with the introduction of every new EOS, its properties are analyzed and the validity of each projection model is investigated for that EOS.

For the imaging scenario, two major families of models exist: physical (rigorous) models and empirical models. The physical models express the physical reality in the imaging process, while empirical models fit pre-defined models to measurements (image data and GCPs) (*Toutin 2004*).

1.8.1 Rigorous / Physical Models

Physical models make use of the metadata (*ephemeris*) provided by the satellite to represent the imaging geometry.

Rigorous sensor models are based on the orbital parameters explained in Section 2.1.2. The most important step in rigorous modeling is to determine the satellite position and the orientation accurately. The actual (observed) ground point's coordinates are then computed (reconstructed)

using the colinearity equations (Section 1.5.2) and the coplanarity conditions (Section 1.5.3). Therefore, the main issue in the rigorous modeling is to obtain an accurate physical model. Ideally, GCPs are not required when the satellite model parameters are obtained perfectly. In practice, *ephemeris* contains measurement errors. Thus, the accuracy of the physical model can be improved with GCPs.

Various rigorous satellite models are proposed in the literature. A representative subset of these models is presented in Chapter 2. Broader literature is covered in *Toutin 2001*, *Dowman 2003*, *Poli 2004a* and *Toutin 2004*.

1.8.2 Empirical Models

Some image providers prefer to deliver empirical model parameters (polynomial coefficients) that are derived from the hidden physical model and provide very fine approximation to the physical models (*Hartley 1997a*, *Grodecki 2001*).

Empirical models use explicit projection functions to represent the imaging scenario. The general form of the projection functions is given by divisions of polynomials that are defined in the object domain:

$$u = \frac{P_1(X, Y, Z)}{P_2(X, Y, Z)} \quad , \quad v = \frac{P_3(X, Y, Z)}{P_4(X, Y, Z)} \quad (1.9)$$

where; u and v are the vertical and horizontal coordinates in the image domain, X, Y, Z are the object domain (geodetic) coordinates, and P_i 's are polynomials. The equations in (1.9) are also referred to as the *forward functions* since they define the projective relation in the forward direction (from the object domain to the image domain).

Different choices for the four polynomials result in different empirical models. Note that, a model with four first order polynomials with $P_2(X,Y,Z)=P_4(X,Y,Z)$ does not correspond to the pinhole camera model since X,Y,Z is defined in a geodetic coordinate system which is not Cartesian. Still, if the region of interest is small enough to approximate the reference ellipsoid as a plane in that region (where the latitude-longitude grid corresponds to a Cartesian grid), the pinhole model can be assumed.

The model with four separate third order polynomials is known to represent the physical reality almost exactly and is known as the Rational Function Model (RFM) (*Grodecki 2001*). There are also simplified empirical models, such as the Linear Pushbroom Camera Model (LPCM) (*Gupta 1997*), Direct Linear Transformation (DLT) (*Wang 1999, Fraser 2001*) and the Affine Projection Model (APM) (*Fraser 2001, Fraser 2004*). These models are explained in Chapter 2.

1.8.3 Registering LR DSMs to the Satellite Images

LR DSMs are frequently used in remote sensing applications. The main usage area is orthorectification, in which the satellite images are resampled on a latitude-longitude or UTM grid (*Leprince 2007*). Another usage is in DEM extraction, where the LR DSM is used either as the source of the GCPs (*Gonçalves 2008*), or as the reference (ground truth) data for accuracy evaluation (*Tsutsui 2007, Jacobsen 2008*). Using an LR DSM as the ground truth may be questionable, especially if the accuracy of the reference LR DSM is close to or worse than the evaluated DEM, which is generally the case for modern EOSs. Regardless of the usage, an LR DSM must be registered to at least one of the stereo satellite image pair, if used.

Registration can be performed by first generating a coarse DEM or a sparse reconstruction from the image pair and then finding the translation to align these (reconstructed) sparse points to the reference LR DSM by using the correlation coefficient (*Gonçalves 2008*). Methods for registering an LR DSM to a single image were non-existent until this study (*Ataseven 2010a*).

1.8.4 Sparse Reconstruction

Sparse reconstruction is the phase that generates the DEM values for a sparse point set of the reference satellite image. Classically, two phases exist; *i*) interest point (feature) detection and *ii*) reconstruction.

Images are often passed through pre-processing before feature detection. The main aims in preprocessing are as follows: *i*) suppressing the noise *ii*) enhancing the features of interest in low contrast regions (e.g. shadows) and *iii*) reducing the radiometric inequalities between the stereo images. The most widely cited preprocessing method is the Wallis filter, which is a basic non-linear filter that forces the local mean and variance to have certain desired values with some forcing parameter (*Wallis 1974, Jayazeri 2008*). Usually the Wallis filter output is used only for feature detection. Matching is achieved by the original images.

The most common feature detector used in the photogrammetry literature is the Förstner detector (*Förstner 1986, Förstner 1987, Jazayeri 2008*), which is a slightly modified version of the Harris feature detector (*Harris 1988*).

Template matching is the most widely used sparse correspondence generation tool in the remote sensing literature. MSE, cross-correlation and their derivatives are used as the matching criteria in many studies (*Jacobsen 2004, Poli 2004a*). There is a special emphasis on one particular algorithm,

namely the *Iterative Least Squares* (ILS), which optimizes a cost function that is computed using both geometric and radiometric errors (*Gruen 1985*). The computational expense of template-matching is typically reduced using a hierarchical matching on image pyramids (*Zhang 2004a, Zhang 2005*).

Edge matching is also used for stereo correspondence generation. Usage of edges provides an improvement in the DSM accuracy and sharpness, especially in the urban regions with tall buildings, where interest point matching is difficult due to occlusions (*Gruen 2005*). Edge detection is generally not the critical step and is performed with one of the well-known algorithms. For matching, epipolar constraints and template matching variants are used (*Zhang 2005*). Edges are more useful in aerial images rather than their satellite counterparts.

For uniform regions, grid based matching can be used. With this approach, matching is possible even for the regions with small spatial intensity variance. Such an approach can be considered as enforcing the geometric consistency constraint rather than intensity based matching since the intensity distribution provides insufficient information for successful matches (*Poli 2004a, Zhang 2005*).

Reconstruction (triangulation) with the rigorous geometric models is performed directly by the collinearity equations under the coplanarity constraint (*Jacobsen 1997*).

For reconstruction using the empirical models, an optimization problem is defined and solved for the optimum world coordinates. The cost function is defined either in the image domain using the forward projection models or in the object domain using the inverse projection models (*Di 2001, Tao 2002*). Image domain cost functions are known to provide better reconstruction accuracy (*Tao 2002*). Classic gradient-descent schemes with

quite coarse initial estimates are preferred for optimization. Due to the analytical form of the empirical models, the gradients can be computed analytically (*Di 2001*).

1.8.5 Dense DSM Generation

Although the ultimate goal of the DSM generation efforts is obtaining dense DSMs, the efforts in photogrammetry are mainly focused on projection model development and obtaining sufficient accuracy in sparse control points. Interpolation is generally presented for completeness. Dense reconstruction without a preceding sparse reconstruction step is uncommon and studies focused on interpolation are few (*Otto 1988*).

The most widely used approach for dense DSM generation is *region growing* which extends the ILS algorithm (*Otto 1988*). The sparse matching results are generally used as the seed points for region growing (*Zhang 2005*). Starting from the seed points; dense correspondences are generated in the neighborhood using the affine transformation of the seed point as the initial estimate. This method is very accurate; but extremely expensive since for every pixel in the image, its correspondence is searched with the ILS. Despite its computational complexity, region growing is the most frequently used technique in dense DSM generation (*Otto 1988, Poli 2007*); possibly since it is already implemented in the commercial software packages.

1.9 Objective of the Study

This study aims at:

1. Developing a fully automatic, GCP-free, dense DSM generation scheme from high resolution stereo satellite images,

2. Reducing the overall computation time for DSM generation.

Airborne imagery is excluded in the study due to the lack of data. Yet, the methods proposed in the thesis study are –in principle- applicable to the aerial imagery with proper adjustments, related to the projection functions.

1.10 Significance of the Study

Currently, DSM generation process requires human intervention in various phases. The most time and resource consuming stage is the GCP collection phase, which requires site visits with expensive measurement equipment. GCP-free methods suffer reconstruction accuracy due to errors (biases) in the projection functions.

The significance of this study is summarized as follows:

1. A fully automatic, GCP-free DSM generation scheme is developed and implemented.
2. A novel, GCP-free method for registering LR DSMs to satellite images is proposed. With this approach, geodetic coordinate estimates are generated for every pixel in the satellite image. The method can be used for all types of map data.
3. A novel, GCP-free bias reduction method for the RPF model is developed. The method combines the registered raster map and sparse optical flow.
4. It is shown that, by proper map registration, simple projective models gain significant accuracy improvement, even when GCPs are absent. Dividing the satellite images into row chunks provide further improvement in the accuracy.

5. A two-step sparse stereo pair generation procedure is developed. The procedure reduces computational complexity while maintaining the matching accuracy without any significant sacrifice in the number of successful matches.
6. Two novel, sparse-to-dense interpolation methods are presented:
 - i) A segmentation-based, edge-preserving algorithm
 - ii) An algorithm that uses integral images and generates a smooth DSM
7. The overall DSM generation time is significantly reduced compared to commercial applications.

1.11 Outline of the Thesis

The outline of the study and its relation to the thesis chapters are shown in Figure 1.9. More detail is presented in Chapter 3 (see Figure 3.1).

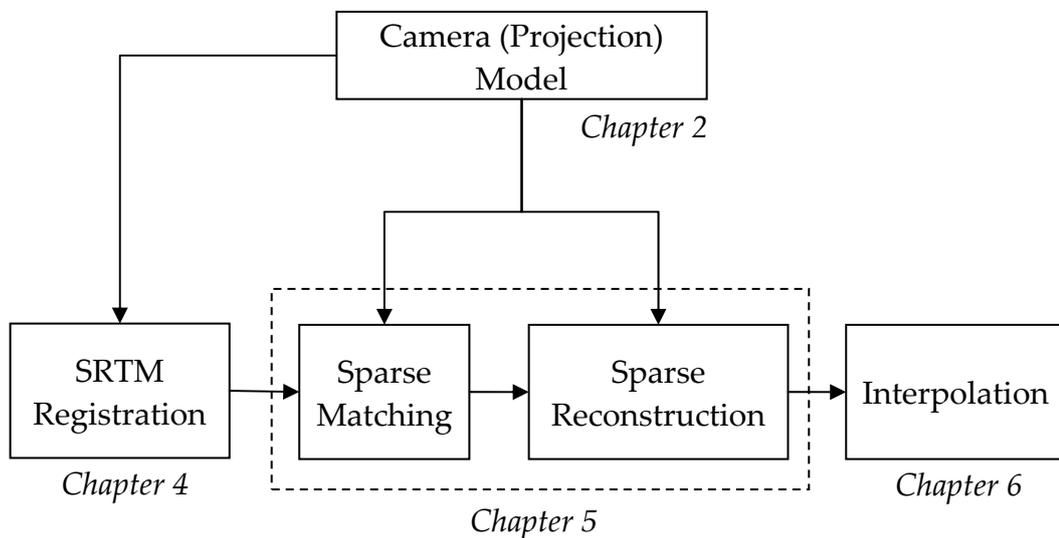


Figure 1.9. Outline of the workflow and its relation to the thesis chapters

In the following sections of the text, the following outline will be followed:

Chapter 2: Projection models for the satellite images are explained.

Chapter 3: The methodology used for DSM generation is introduced.

Chapter 4: The methodology for Registration of SRTM data and satellite images is described.

Chapter 5: Sparse reconstruction strategy is explained.

Chapter 6: Dense DSM generation via interpolation with four different approaches is described.

Chapter 7: Experiments are presented and the results are reported.

Chapter 8: The thesis is concluded.

Appendix A: Usage of the SRTM registration in estimating the parameters of simpler empirical models is presented.

Appendix B: The details of RPC reconstruction are presented.

CHAPTER 2

PROJECTION MODELS

2.1 Rigorous Models

2.1.1 Kepler's Laws

All satellites obey the Kepler's Laws on orbital dynamics. Kepler's laws can be summarized as:

1st Law: All satellites move on an elliptical orbit, and The Earth's center of gravity lies at one of the elliptic foci.

2nd Law: The vector from Earth's center to the satellite sweeps equal area in equal time periods.

3rd Law: The satellite's rotation period is given by $P^2 = \frac{4\pi^2 a^3}{GM_e}$, where a

is the semi-major axis of the orbital ellipse, G is the gravitational constant ($G = 6.67428 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ (*Mohr 2006*)), and M_e is the mass of The Earth ($M_e = 5.9736 \times 10^{24} \text{ kg}$ (*Yoder 1995*)).

Kepler also invented methods for conversions and methods to compute satellite position at a given time (which do not have explicit solutions).

2.1.2 Orbital Parameters

Orbital parameters of the satellite motion are used in rigorous physical models of the image projection scenario. Even if simpler models can be used for the imaging scenario, it is required to have an understanding for the basic parameters of satellite motion in orbit. In Figure 2.1, a general diagram of satellite orbit is presented with the main orbital parameters. The parameters displayed on this image are defined as follows:

Perigee: The point in the orbit where the satellite is closest to The Earth's center of gravity

True anomaly: The angle between the two rays: Earth center-perigee (EP), and Earth center-satellite (ES). The angle is measured from EP to ES, in the satellite's movement direction.

Ascending node (λ_{AN}): The geodetic longitude that the satellite passes the equatorial plane towards North. Due to Earth's rotation, λ_{AN} is different for every tour (track) of the satellite.

Descending node (λ_{DN}): The geodetic longitude that the satellite passes the equatorial plane towards South. Due to Earth's rotation, λ_{DN} is different for every tour (track) of the satellite.

Semi-major axis (a) : Half of the major axis length of the orbital ellipse (see Figure 2.2).

Semi-minor axis (b): Half of the minor axis length of the orbital ellipse (see Figure 2.2).

Orbital Eccentricity (e): A value that gives how much the orbital ellipse deviates from a circle.

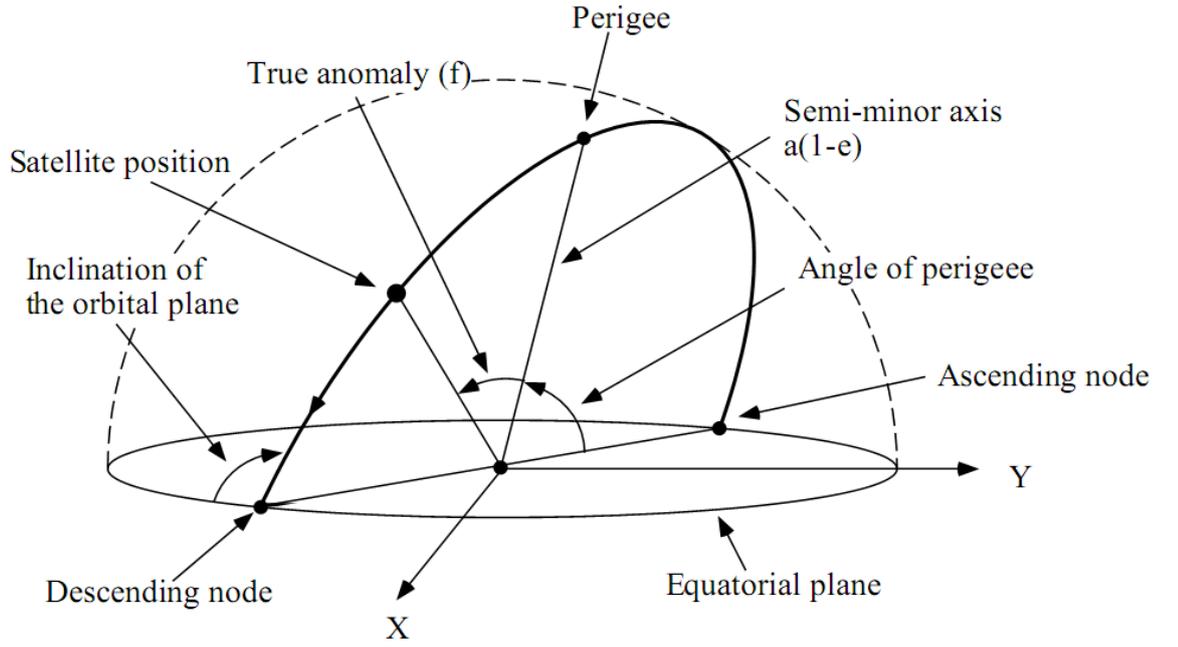


Figure 2.1. A Satellite and its orbital parameters (taken from Gupta 1995).

It should be noted that, due to the earth's rotation, the difference $\lambda_{DN} - \lambda_{AN}$ is not equal to 180 degrees. The relations for λ_{DN} and λ_{AN} are (Gupta 1995):

$$\begin{aligned}
 \lambda_{DN_k} &= \lambda_{DN_1} - (k-1) \times T(360^\circ) \\
 \lambda_{AN_k} &= \lambda_{DN_k} + 180^\circ + w_e \times T(180^\circ) \\
 \lambda_{T_k} &= \lambda_{AN_k} - 90^\circ + w_e \times T(90^\circ)
 \end{aligned} \tag{2.1}$$

Here, λ_{T_k} is the perigee longitude for the k^{th} track. $T(x)$ is the time that is required for the satellite to build an angle of x degrees from the ascending node. $T(x)$ is a fixed function for all tracks, since it is dependent only on the satellite orbit. w_e is the earth's angular rotation speed ($w_e = 7.292115 \times 10^{-5}$

rad.s⁻¹ (Yoder 1995)). As seen in (2.1), both ascending and descending nodes for the k^{th} track can be obtained directly from the previous track's values.

The relation between a , b and e is:

$$b = a\sqrt{1 - e^2} \quad (2.2)$$

For determining actual satellite position in the orbit, a *reference circle* is used. The reference circle lies in the orbital plane and its radius is the semi-major axis of the orbit. The reference circle is shown in Figure 2.2.

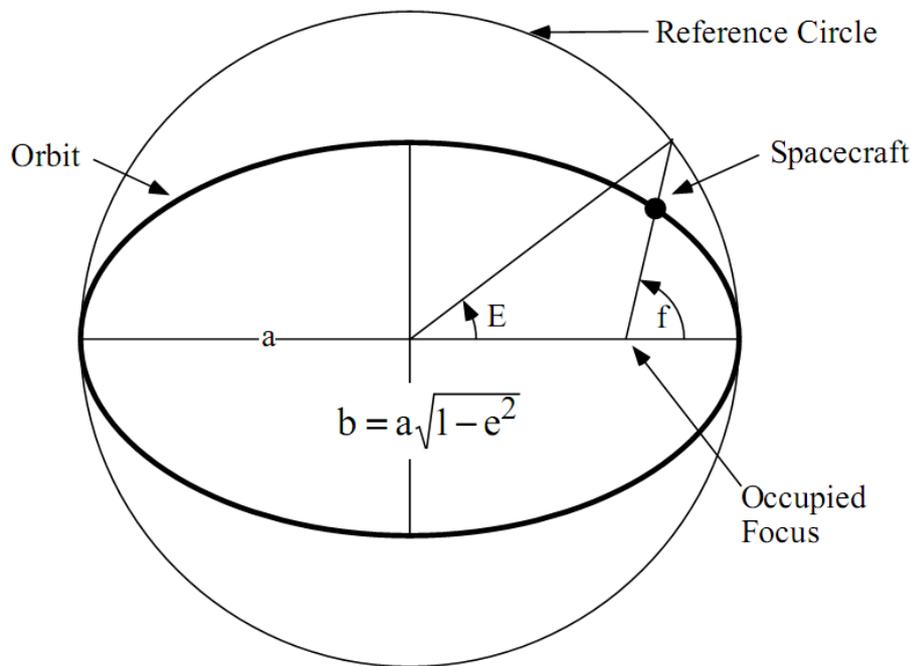


Figure 2.2. Satellite orbit and the reference circle. (taken from Gupta 1995)

In Figure 2.2, E (called the *eccentric anomaly*) refers to the central angle of the reference circle, for the point obtained by intersecting the “Earth center - satellite position line” (the associated vector is referred to as “*satellite*

vector”) with the reference circle. There is also an intermediate term called *mean anomaly*, denoted by M , which is used in finding satellite position.

The EOSs sense the time they pass from the perigee and measure the time elapsed from the last perigee visit to the instant that the center line of the acquired image is taken. The procedure to determine the satellite position, given the elapsed time (t) from the last perigee visit is as follows:

i) From Kepler’s 3rd law, obtain the mean angular velocity:

$$\bar{w} = 2\pi / P = \sqrt{\frac{G.M_e}{a^3}} \quad (2.3)$$

ii) Obtain mean anomaly:

$$M = t \times \bar{w} \quad (2.4)$$

iii) Solve

$$M = E - e \cdot \sin(E) \quad (2.5)$$

iteratively for E (use $E_0 = M$ as initial value). Since the orbits are very close to circles, M and E are similar and the iteration terminates in a few steps (*Gupta 1995*)

iv) Obtain *true anomaly*:

$$f = \tan^{-1} \left(\frac{\sqrt{1-e^2} \sin E}{\cos E - e} \right) \quad (2.6)$$

v) From the *true anomaly* and λ_{AN_k} for that track, obtain the geocentric, geo-fixed longitude.

$$\lambda_f = \lambda_{AN_k} - (90^\circ + f) + w_e \times T(90^\circ + f) \quad (2.7)$$

vi) From the true *anomaly* and orbital inclination (θ), obtain the **geocentric** latitude of the satellite. This can be obtained by applying a rotation to the perigee vector (EP) about the axis perpendicular to the orbital plane, by an angle of f , and taking the resultant polar angle.

vii) Convert the geocentric coordinates (angles) of the satellite to geodetic counterparts using (1.5).

The obtained latitude and longitude values should give the ground point just below the satellite. The satellites are designed in such a way that their look direction is perpendicular to the ground point (*nadir* point) they fly above. Therefore, the satellite geodetic angles are the same with the ground point they are directed to.

Note that this formulation is for the nadir image alone. For the *aft* and *fore* imaging cases, the computations are the same for the satellite position, whereas more complicated for the imaged point.

2.1.3 The Geometrical Models

The mainstream in rigorous modeling can be outlined as follows: The sensor model parameters are divided into two groups; exterior orientation parameters and error correction parameters. The exterior orientation (coarse) parameters define the satellite position and orientation (*attitude*). Error correction (fine) parameters are used for modeling the non-idealities (variations/drifts) of the imaging system and the imaging scenario. Most of the time, some of the used parameters are highly correlated and calibration may not result in actual (physical) parameters even if the generated DEMs

are accurate. Due to this fact, the satellite models can be generated using the most dominant parameters. However, these correlations are satellite dependent and cannot be generalized, and consequently, “a generic satellite model” is difficult to achieve. Still, there are sensor models that are applicable to more than one satellite.

A model was proposed by Gagan and Dowman (1988). Poli’s explanation on this model and the clues given in Dowman’s recent publications are as follows:

Gagan and Dowman proposed a dynamic orbital parameter model (Gagan 1988). The satellite movement along the path is described by two orbital parameters (true anomaly and the right ascension of the ascending node), that are modeled with linear angular changes with time and included in the collinearity equations. The attitude variations are modeled by drift rates (*Poli 2004a*).

Dowman’s model uses second order polynomials of time for modeling the positions of the sensor array, while it acquires successive image lines. For 3D to 2D mapping, the rigorous collinearity equations given in (1.12) are used. Later, Dowman and Michalis extended the model to a generic rigorous sensor model (with 14 parameters for the exterior orientation of the satellite) that are applicable to today’s popular satellites, and reported RMS errors less than 2 pixels for ASTER data with 3 GCPs (*Dowman 2003*). Then, with the same model, they reported DEM accuracy better than 10m (except steep regions) for SPOT HRS data. (*Michalis 2004*)

Gupta and Hartley proposed an application of Levenberg-Marquardt algorithm (*Hartley 1993*) to obtain optimum satellite parameters (*Gupta 1995*). They try to minimize RMS re-projection error between the calculated and measured image points of 25 GCPs. The main approach is to determine a coarse update for satellite position and orientation, and then refine it using

the *attitude drift* parameters. Their parameter vector contains: Semi-major axis a , eccentricity e , inclination i , perigee angle w , longitude of the DN, look angle Ψ_{x1} , look angle Ψ_{y1} , look angle Ψ_{xn} , look angle Ψ_{yn} time perigee to current position, and *dwelt time*. They report sub-pixel accuracy for 90% of the pixels of a SPOT stereo pair (Hartley 1995).

Poli (Poli 2003) divides the image into successive chunks of image lines, and models external orientation (position and attitude) change between these chunks (during acquisition) as a second order polynomial of time, with continuity constraints up to the second order derivative. The original paper (Poli 2003) states that the continuity constraints reduce the external orientation polynomials to the first order. For the error correction, Poli uses principal point displacement ($\Delta x_p, \Delta y_p$), focal length variation (Δf), the symmetric (k_1, k_2) and decentering (p_1, p_2) lens distortion, the shear factor in y direction (s_y), and CCD line rotation in the image plane (θ) for modeling systematic errors. This method resulted in 8m planimetric, and 10.4m elevation error (RMS) for ASTER images and 6 GCPs, but the error was measured only at the GCPs (Poli 2004b). The author reports that the experiment on SPOT HRS data resulted in planimetric errors less than 7m and elevation error less than 5m if 8 or more GCPs are available (Poli 2003). The author also reports that the error correction parameters, k_1, k_2, p_2, s_y , and θ are effective on the accuracy.

Jacobsen published another model (Jacobsen 1997), which included parameters for external orientation, focal length, lens distortions (fish eye effect), earth curvature, atmospheric refraction, etc. This model is applied to several satellites successfully (see Poli 2004a for the references).

The actual physical models that are generated by the satellite designers are rarely public. In fact, the usability of these models is questionable for the

end user, since these detailed models for the new generation satellites are extremely complicated (Grodecki 2001). Thus, the current trend in satellite photogrammetry is dominated by the empirical models.

2.2 Empirical Models

2.2.1 Rational Polynomial Function Model (RFM)

Rational polynomial function model aims at providing an accurate empirical model that has a generic form. The main idea is a direct extension of the pinhole camera relations. For a pinhole camera, a 3D object point is mapped to a 2D image point by a 3x4 projection matrix with the relation:

$$\begin{bmatrix} ku \\ kv \\ k \end{bmatrix} = \begin{bmatrix} \bar{P}_1^T \\ \bar{P}_2^T \\ \bar{P}_3^T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.8)$$

where; $\bar{P}_i^T P$'s are the 1x4 row vectors of the projection matrix; X , Y and Z are the real world (object) coordinates; u and v are the image domain coordinates of the projection. Obviously, u and v can easily be obtained by dividing the product by k , which is given by $\bar{P}_3^T [X Y Z]^T$. The inner products of the projection matrix rows and the position vector are in fact the first order polynomials of the object domain coordinates. Consequently, the image coordinates (u,v) can be found by the division of two first order polynomials of the object coordinates.

For the pushbroom sensors case, the geometry is more complicated (different focal point for each image line, attitude changes, etc.) and therefore, higher order polynomials are required for accurate modeling. This

is the underlying idea of the RFM. Rational Polynomial Coefficients (RPCs) are simply the coefficients of the polynomials in the RFM. The most widely used RFM is the third order model, which is shown to represent the physical situation almost exactly (*Hartley 1997a, Fraser 2001, Grodecki 2001, Dial 2002a, Hu 2004, Hosseini 2008*):

$$y = \frac{P_1(X,Y,Z)}{P_2(X,Y,Z)} \quad , \quad x = \frac{P_3(X,Y,Z)}{P_4(X,Y,Z)} \quad (2.9)$$

Here, P_1 , P_2 , P_3 and P_4 are third order polynomials of the three variables X , Y and Z . Apart from the pinhole convention, X , Y and Z are not defined in the Cartesian coordinates. Instead, geodetic latitude, longitude and altitude (height) are used. For better accuracy, the geodetic and image coordinates are normalized to $[-1 \ 1]$ interval in the RFM relations (*Hartley 1997a, Grodecki 2001, Tao 2002*). For notational consistency, we state the following definitions:

$$X = \frac{\phi - \phi_0}{\phi_{scale}} \quad , \quad Y = \frac{\lambda - \lambda_0}{\lambda_{scale}} \quad , \quad Z = \frac{h - h_0}{h_{scale}} \quad (2.10)$$

In (2.10), ϕ_0 , λ_0 and h_0 represent the mean latitude, longitude and height values of the observed region, respectively. The scale factors in the denominators are used to map the object coordinates onto the $[-1,1]$ interval. The image domain coordinates are also normalized similarly:

$$y = \frac{u - u_0}{u_{scale}} \quad , \quad x = \frac{v - v_0}{v_{scale}} \quad (2.11)$$

This way, the projection functions are determined such that the projections are always in the $[-1,1]$ interval. Thus, corresponding de-

normalization is performed after the projection to obtain the image coordinates. Normalization improves the accuracy of the projection functions for finite numerical precision implementations. The normalization parameters are given together with the RPCs by the image providers.

There is no known method to extract physical parameters from given RPCs, since many physical parameters are highly correlated. Image providers use the physical model to generate accurate RPCs through simulated projections (*Grodecki 2001*). Since the RPCs are scene-dependent, users need new RPCs for every new image. Obviously, in theory, computation of 3rd order RPF coefficients using GCPs is possible. However, a third order, three-parameter polynomial has 20 coefficients; and the RFM has 4 separate polynomials. Thus, this task requires at least 40 GCPs per image. Furthermore, the coefficient determination problem is ill-posed. For accurate projection functions, well-distributed GCPs are required (*Hosseini 2008*). Thus, an effort to determine RPCs using GCPs without any physical model should be avoided (*Jacobsen 2008*).

The main advantage of RPFs is due to their quite fine approximations to the physical model; the end user does not require the complicated physical model (*Dial 2002b*).

Block Adjustment

RPF models may be improved further by *block adjustment*, in which the errors are corrected in the image domain by correction functions with a least squares approach (*Dial 2002a, Grodecki 2003, Gruen 2005*). This is done using GCPs as the ground truth. For the stereo or multi-view case, additional image domain correspondences (without ground control data) are also useful (*tie points*).

In block adjustment, an error correction function is assumed (for each block) and its parameters are optimized to fit the GCPs and the tie points.

It has been shown that the error in RPF can be corrected accurately by a bias correction in the image domain (*Dial 2002a*), at least for IKONOS and Quickbird images. This conclusion is expected to be valid for narrow FOV cameras, whose characteristics are close to the affine projection. The errors in the projection information appear to be caused by invalid information for the initial position and the orientation of the camera (long term error accumulation) and remain mostly unchanged for the short imaging period (i.e. short term error accumulation is small).

Inverse Functions

The projection functions defined by the equation (2.9) are often named as *forward functions*. There is another set of equations (*inverse functions*) which represent X and Y as functions of u , v and Z :

$$X = \frac{P_5(u, v, Z)}{P_6(u, v, Z)} \quad , \quad Y = \frac{P_7(u, v, Z)}{P_8(u, v, Z)} \quad (2.12)$$

The inverse equations can be obtained easily from the forward equations using simulated projections and solving the matrix equation for the coefficients of P_5 , P_6 , P_7 and P_8 . The advantage of these functions is that, once the Z value is obtained, X and Y are instantly available for each pixel. Thus, the optimization problem for DEM generation reduces to 1D search (*Di 2001*). However, in the reconstruction with the inverse functions, the cost function is defined in the object domain and the results are not as accurate as the solutions for the forward functions (*Tao 2002*).

Another usage of the inverse functions is that, *epipolar curves* can be easily obtained by using (2.12) for the first image and (2.9) for the second image. For an interest point (u_1, v_1) in the first image; different X,Y values are obtained by changing the Z value in (2.12). For each resultant (X,Y,Z) , the projection is computed in the second image using (2.9). Combining the projected points (for different Z) values in the second image, an epipolar curve is obtained (*Wang 2010*). One direct result of this strategy is that, once the epipolar curve is obtained, the match point (u_1, v_1) in the second image (that lies on the epipolar curve) gives the 3D reconstruction instantly, since the (X,Y,Z) is known everywhere on the epipolar curve. However, this approach is not numerically stable, since small errors in the forward functions affect the inverse functions and the resultant epipolar curve may not pass through the actual match point. Thus, for this approach to be successful, the rational projection functions must be corrected via block adjustment by using GCPs (before computing the inverse function parameters).

2.2.2 Direct Linear Transformation (DLT)

Direct linear transformation (*Wang 1999, Fraser 2001*) is a simpler form of RPFs, as first order polynomials and a single second order term, which accounts for the curved nature of satellite imaging (mainly due to orbit). The DLT equations are given as (*Wang 1999*):

$$v = \frac{L_1X + L_2Y + L_3Z + L_4}{L_9X + L_{10}Y + L_{11}Z + 1} \tag{2.13}$$

$$u = \frac{L_5X + L_6Y + L_7Z + L_8}{L_9X + L_{10}Y + L_{11}Z + 1} + L_{12}uv$$

In these equations; u , v , X , Y and Z are the normalized coordinates as explained in the RFM section.

In (2.13), DLT adds a second order term in the flight direction of the satellite to handle the orbital movement of the optical center. Note that the first equation is not the same with the pinhole case, since the X , Y and Z values are originated from the geodetic coordinate system, which is not Cartesian. It should also be noted that, triangulation with DLT is the same with that of the RPF model.

DLT is reported to provide sufficient accuracy for SPOT and IKONOS image pairs (*Wang 1999, Fraser 2001*). However, there is an inconsistency on the DLT formulation among authors. In the original paper by Wang, and in the Fraser's work, the second order term is in the u equation. Jacobsen omits the second order term, and converts the unities in the denominator to constants, obtaining the pinhole model (but still denotes it as DLT) (*Jacobsen 2008*).

Experimental results on DLT show that, if more than 8 GCPs are used for parameter estimation, DLT is as accurate as higher order RPFs (*Wang 1999, Fraser 2001, Jacobsen 2008*).

2.2.3 Linear Pushbroom Camera Model (LPCM)

On the other hand, Gupta and Hartley, who are researchers from vision community, proposed a simple approximation to the physical model, which results in generic equations (*Gupta 1997*). Their model assumes the following:

- The satellites moves in a straight line with constant velocity during acquisition
- The sensor attitude is fixed during acquisition

This model can easily be converted to a closed form projection equation:

$$\begin{bmatrix} u \\ kv \\ k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & f & p_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/V_x & 0 & 0 \\ -V_y/V_x & 1 & 0 \\ -V_z/V_x & 0 & 0 \end{bmatrix} (\mathbf{R} \mid -\mathbf{RT}) \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2.14)$$

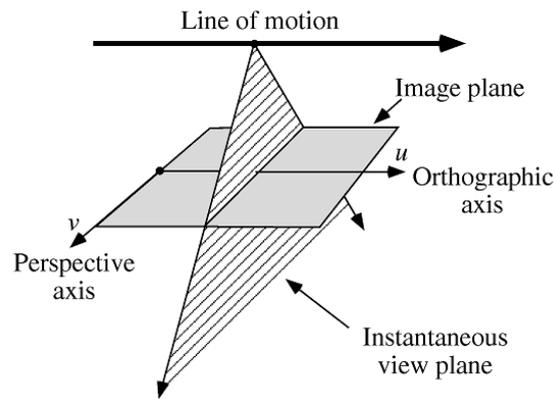


Figure 2.3 Linear pushbroom camera model. Taken from *Gupta 1997*

Here, u, v are the image coordinates and x, y, z are the (Cartesian) world coordinates of the observed point, k is a scale factor, V_x, V_y and V_z are components of the relative velocity of the satellite w.r.t. the earth surface beneath it, f is the focal length, p_v is the position of the center pixel, \mathbf{R} is the rotation matrix and \mathbf{T} is the translation vector between the orthogonal image and the world coordinate axes. As the equation suggests, the LPCM has 11 parameters.

Gupta and Hartley further investigate the model and show that a stereo image pair has a fundamental matrix \mathbf{F} , which satisfies the following relation:

$$[u' \quad u'v' \quad v' \quad 1] \mathbf{F} \begin{bmatrix} u \\ uv \\ v \\ 1 \end{bmatrix} = 0 \quad (2.15)$$

\mathbf{F} is a 4x4 matrix whose top-left 2x2 block equals to zero. For a point (u, v) in one image, this relation defines a hyperbola (epipolar curve) on the second image and its conjugate point lies on this curve. Obviously, equation (2.14) holds for any multiple of \mathbf{F} , and therefore, \mathbf{F} has 11 independent parameters which can be solved by 11 correspondences. Gupta and Hartley use the popular 8-point algorithm (*Hartley 1995*) for numerical stability.

They also describe a linear method to determine the parameters of the 4x3 projection matrices from \mathbf{F} , up to an affine transformation of space, and end up with a matrix equation, whose solution is the desired 3D point, again up to an affine transformation in 3D space. Thus, with at least 4 GCPs, the affine transformation that maps the actual 3D points to the reconstructed scene can be determined. The actual coordinates can then be easily obtained by its inverse.

The assumptions for LPCM neglect the effects of the orientation and distortion parameters used in all rigorous modeling efforts. However, the experimental results show that the requirement for extra care in order to determine the satellite parameters is questionable (*Gupta 1997, Lee 2003*). Although the results in these papers show 3 times larger errors compared to rigorous models, the accuracy depends mainly on the resolution, and one can anticipate that DEM error below 3m is possible with the resolution of modern satellites. Thus, this model has the potential to provide sufficient DEM accuracy for many applications.

2.2.4 Affine Projection Model (APM)

Affine Projection Model is a highly simplified model. In fact, it is a DLT with unity denominators and no second order term (8 parameters):

$$\begin{aligned}u &= a_1 + a_2X + a_3Y + a_4Z \\v &= a_5 + a_6X + a_7Y + a_8Z\end{aligned}\tag{2.16}$$

There is also an extended version of the APM, namely the 3D Affine Transformation.

$$\begin{aligned}u &= a_1 + a_2X + a_3Y + a_4Z + a_9XZ + a_{10}YZ + a_{13}X^2 \\v &= a_5 + a_6X + a_7Y + a_8Z + a_{11}XZ + a_{12}YZ + a_{14}XY\end{aligned}\tag{2.17}$$

The idea in APM is that, due to the high focal length of the satellite imaging devices, the incident light reaching the sensors are almost parallel. Therefore, an affine camera model can be an approximation to the satellite imaging system. (*Fraser 2001, Fraser 2004*). Fraser's interesting results show that, affine model can perform close to the third order RPF model, especially for the scenes with small height variance. Jacobsen confirms this experimentally for various EOSs and also reports that for more than 8 GCPs, APM is as accurate as the RPF model (*Jacobsen 2008*).

Reconstruction with the affine model is similar to that of the other methods. Hartley's polynomial method (*Hartley 1997b*) is also applicable to the APM if (2.16) is used.

One important step in APM is to transform the image into an affine version, in the perspective direction (cross-track direction) (*Fraser 2001*).

CHAPTER 3

PROPOSED METHODOLOGY

The workflow used in the thesis study is composed of pipelined operations (Figure 3.1). Each block in Figure 3.1 has well-defined inputs and outputs and can be combined with methods other than the ones that are presented in this text. Therefore, the contributions of the study can be considered as a whole or block-wise.

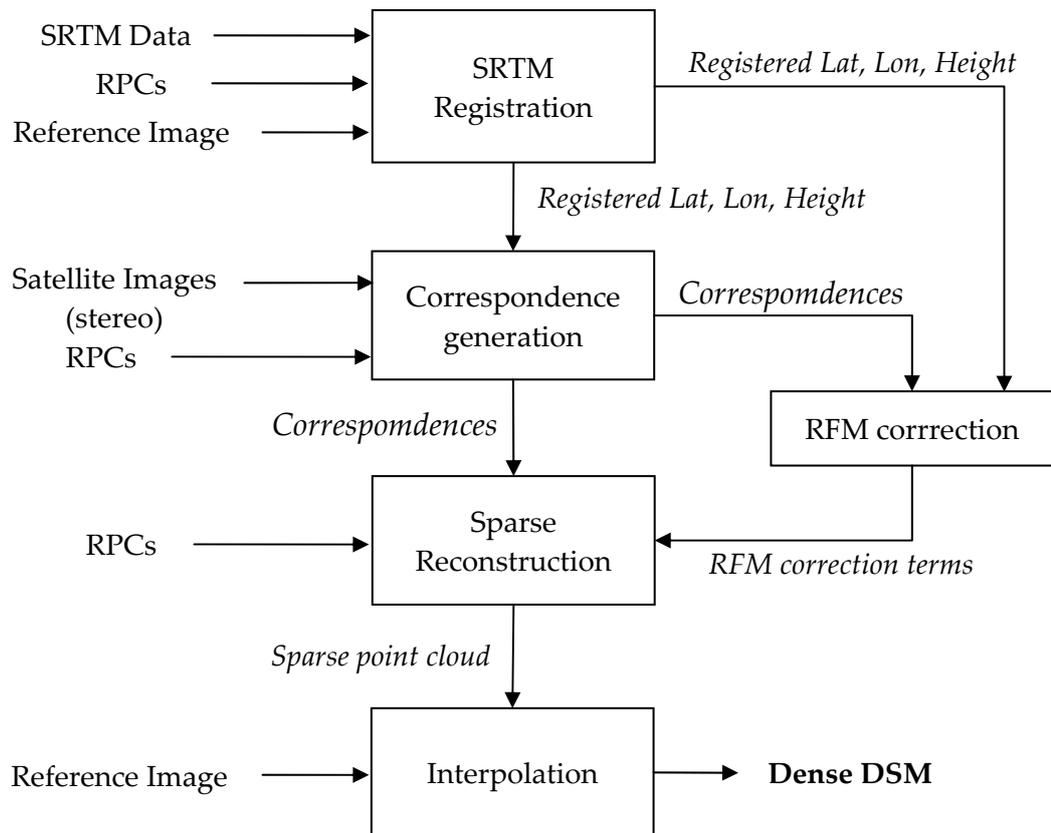


Figure 3.1. Thesis study workflow

Every block in the workflow is developed for a GCP-free process. However, the approaches presented here can well be applied, if GCPs are also available. In such a case, for better accuracy, RFM correction must be performed by the GCPs as the first step and all other blocks should use the GCP-corrected RFMs.

Developing a fully automatic, GCP-free workflow is difficult, since the advantage of having precise projection functions is not available anymore. The projection functions delivered by the image providers typically have five to ten pixels of systematic projection error (bias), which could be corrected by GCPs. The disadvantages of GCP-free operation can be listed as follows:

- i)** Accurate bundle adjustment cannot be performed
- ii)** Epipolar curves cannot be used in the stereo correspondence generation process. Consequently, the search operation for correspondence match is computationally expensive.
- iii)** The DEM reconstruction accuracy is below the potential of the imaging system.

The above mentioned difficulties bring the requirement for specific approaches in GCP-free studies. In fact, studying without reference data undermines the DEM generation process in general and a GCP-free process is not expected to provide the accuracy figures of a GCP-enabled one. Still, having GCP-free workflow is valuable, when GCPs are not available or cannot be obtained due to different reasons.

3.1 SRTM Registration

Our approach uses less accurate, but freely available LR-DSMs to obtain “anchors” in the building blocks. Public raster DSMs, such as ASTER GDEM and SRTM, have limited accuracy both in geolocation and height in general. However, the accuracy of the LR-DSMs obtained from space shuttles is not uniform everywhere. That is, for some ground locations, the LR-DSMs are quite accurate while for some others the accuracy is poor. For instance, SRTM is obtained with radar interferometry. As the analysis suggest, the elevation accuracy of the SRTM is very good (less than 2m) in flat regions (except for sand and water bodies) and worse in mountainous terrain (*Rodriguez 2005*).

One advantage of LR-DSMs is the following: Since an LR-DSM (e.g. SRTM) gives a *height function* of latitude and longitude (in the form $h(\phi, \lambda)$), the geolocation error is ineffective for flat regions. A geolocation error $(\Delta\phi, \Delta\lambda)$ means that, for the location (ϕ, λ) , SRTM has the height value for $h(\phi + \Delta\phi, \lambda + \Delta\lambda)$. By definition, for the flat regions, $h(\phi, \lambda) \approx h(\phi + \Delta\phi, \lambda + \Delta\lambda)$. Therefore, even if there is an overall geolocation bias in the SRTM, it is not effective for a smooth terrain. In other words, if one asks the SRTM for the height value at the location (ϕ, λ) , it will return the correct value if (ϕ, λ) lies in a flat region, where the SRTM data is accurate.

Still, SRTM cannot replace GCPs, even for the flat regions, since the SRTM points are not marked in the satellite images. Thus, one must determine a technique to utilize its accuracy by establishing a method to “connect” (i.e. register) the SRTM points to the satellite image pixels.

3.2 Sparse Correspondence Generation

For sparse correspondence generation, in the absence of epipolar curves, a GCP-free approach should use a method to narrow the search region.

One option is based on the following strategy: *i)* compute the inverse RFMs from the biased forward RFMs; *ii)* draw the erroneous epipolar curves using the inverse RFMs and *iii)* perform the search operation in a “epipolar strip” centered on the epipolar curve. Such an approach is flawed by the errors in inverse RFM computation, which is based on the solution of a homogeneous linear system of equations (in the form of $\mathbf{Ax}=\mathbf{0}$) obtained from the forward RFMs and a number of simulated projections. The image domain projection errors (of the biased RFMs) are included into the system matrix (\mathbf{A}) vector, which is then put into the singular value decomposition for the solution. The error analysis of this step is not easy since the properties of the system matrix are dependent on the locations of the selected simulation points and the forward RFM coefficients, which are scene-specific. Thus, in the GCP-free case, if derived inverse RFMs are used, it is not possible –in general- to predict the accuracy of the epipolar curves that will determine the strip width. If a wide strip is used to assure success, the benefit of the epipolar curve will disappear.

Another approach is to use a coarse DSM that is obtained from the triangulation of fewer correspondences which are obtained by a hierarchical search and performing interpolation to fill the empty pixels. This coarse DSM is then used to generate the initial estimates for many sparse correspondences.

The approach used in this study resembles the latter approach, but instead of generating a coarse DSM, LR-DSMs (which are already available) are utilized (i.e. SRTM registration). Generating a coarse DSM is not a better choice for a GCP-free study, since the biased RFMs cannot provide more accurate DSM than the popular LR-DSMs. In fact, since one only needs the initial estimates for the matching algorithms, the coarse DSM or a registered LR-DSM does not need to be very accurate. Using an LR-DSM has the advantage that, the known world coordinates are regularly sampled. Thus, the interpolation scheme is fairly simple.

3.3 RFM Correction

As mentioned earlier, LR-DSMs cannot be fully utilized as GCPs. The approach in this study uses an opportunity due to stereo data, by utilizing the stereo correspondences. We start with the biased projection functions to obtain a biased SRTM registration. Then a few accurate stereo correspondences are generated, preferably on the smooth terrain. The displacement vector between the initial estimate and the final match point is used to equalize the bias in projection functions of the two images. As explained in Chapter 5 and demonstrated in Chapter 7, this technique improves the accuracy in the reconstructed height with a small increase in the geolocation error.

3.4 Sparse Triangulation

Sparse triangulation (reconstruction) is an intermediate step of the workflow and the thesis study does not propose any novelty in this topic.

We use the formulation presented by *Di 2001* and the Levenberg-Marquardt (LM) algorithm for optimization. In the manuscript, Di stops after the derivation of the derivatives and leaves the selection of the optimization algorithm to the reader. We preferred LM, since its convergence characteristics can be controlled by its damping parameter. In fact, this parameter affects the convergence so dramatically that, if the method is used in commercial software, the user can be allowed to select the operating point in the “speed vs reliability” trade-off.

The accuracy in the sparse triangulation step is quite critical for the final dense DSM. As expressed earlier, a GCP-free process is not expected to provide the accuracy of a GCP-enabled one. In the proposed technique in Chapter 4, the geolocation accuracy can be sacrificed for better elevation accuracy by using the output of the RFM correction block. The experiments show that, the sacrifice in the geolocation accuracy is small and the gain in the elevation accuracy is satisfactory.

3.5 Dense DSM Generation

After the sparse reconstruction phase, the interpolation is performed. The current trend in the literature uses a type of region grow algorithm, in which correlation-based matching is performed for each individual pixel. This is an old and expensive method. Our presumption is that, since this method is already implemented in the commercial software, the authors (who contribute mostly in geometrical models) use the software to be able to report a complete study; hence, the publications on sparse-to-dense conversion are very rare.

In this study, three different interpolation methods are proposed (see Chapter 5):

- i) A fast and simple method that uses integral image interpolation using rectangular or circular windows.
- ii) Adoption of a novel method (*Çiğla 2012*) (usage of color-aware integral images of height) for sparse height, with or without pre-interpolation with method (i)
- iii) Adoption of the Speeded-up Turbo Pixels (*Çiğla 2010*) segmentation algorithm to dense DSM generation.

CHAPTER 4

SRTM REGISTRATION

4.1 Available Digital Surface Models

4.1.1 SRTM

The SRTM grid is a geodetic (latitude-longitude) grid and the sampling is accordingly. The sampling interval is 1 arcsecond in the USA and 3 arcseconds elsewhere. Such sampling corresponds to 30m and 90m GSD on The Equator. The GSDs vary with the latitude, especially along the longitude axis. However, the GSDs are often mentioned in an ergonomic fashion, as 30m and 90m.

SRTM has voids in the dataset due to the nature of the imaging technology:

Voids were caused by two main mechanisms: steep slopes facing away from the radar (shadowing) or towards the radar (foreshortening or layover) and smooth areas such as smooth water or sand which scattered too little energy back to the radar to create an image (Farr *et al.* 2007).

A very detailed SRTM error analysis is presented by *Rodríguez (2005)*. Here we only quote the general information:

Table 4.1. Height error figures for the SRTM data compared with kinematic GPS transects (Rodríguez 2005).

Continent	Mean (m)	Standard Deviation (m)	Absolute Error (LE90) (m)
Africa	1.3	3.8	6.0
Australia	1.8	3.5	6.0
Eurasia	-0.7	3.7	6.6
North America	0.1	4.0	6.5
New Zealand	1.4	5.9	10.0
South America	1.7	4.1	7.5

The CE90 for geolocation errors are given in Table 4.2. As seen in the error figures, SRTM is very accurate for the majority of points, especially in height. Although geolocation errors are relatively higher, they are mostly below 10m. In fact, since SRTM is a DSM that states the height value for a (latitude, longitude) coordinate, geolocation errors are not effective for smooth regions (see section 1.4). This property of the SRTM data is utilized in this study.

Table 4.2. SRTM geolocation error figures compared with kinematic GPS transects (Rodríguez 2005).

Continent	Mean Abs. Geolocation Error (m)
Africa	11.9
Australia	7.2
Eurasia	8.8
North America	12.6
Islands	9.0
South America	9.0

4.1.2 ASTER GDEM

ASTER GDEM is another low resolution DSM, obtained from the large ASTER image database using automated algorithms. It reflects the top of the land cover due to the source of data (ASTER images). ASTER GDEM Version 2 was released in October 3, 2011 (*J Space Systems 2011*). Validation results indicate that the second version provides significant accuracy improvement compared to its predecessor (*Tachikawa 2011*).

Compared to SRTM data,

- i) ASTER GDEM covers a larger latitude range,
- ii) ASTER GDEM has a smaller GSD outside the USA,
- iii) ASTER GDEM data is generated from more recent data,
- iv) ASTER GDEM is generated using automated algorithms, thus it may have errors caused by the image content,
- v) ASTER GDEM is more reliable for mountainous regions (steep regions), and
- vi) ASTER GDEM has less accuracy.

Despite these promising GSD of ASTER GDEM, the recent studies on the accuracy comparison between the SRTM and the ASTER GDEM has shown that the SRTM data is superior in accuracy. (*Chang 2010, Frey 2011*) The ASTER GDEM documents also state that the higher GSD of ASTER GDEM does not yield better resolution:

Also, while the elevation postings in the ASTER GDEM are at 1 arcsecond, or approximately 30 m, the detail of topographic expression resolvable in the ASTER GDEM appears to be between 100m and 120m (*ASTER GDEM Readme*).

4.1.3 LIDAR / LADAR

LIDAR data are in the form of sparse point clouds and therefore, LIDAR is not an actual DSM. However, airborne LIDAR has very high GSD and therefore, DSM on a regular grid can be obtained by interpolating the LIDAR data.

Spaceborne LIDAR is practical and very accurate elevation measurements are obtainable. The NASA satellite ICESat provided sub-meter elevation accuracy in its active years (*Abshire 2005, Schutz 2005*). However, with spaceborne LIDAR, only irregular and sparse ground samples can be obtained. Thus, the spaceborne LIDAR data can only be used as the reference data for DSM accuracy assessment.

Spaceborne LIDAR on a regular grid with reasonable GSD is unavailable –to this day- due to technological limitations. Therefore, in photogrammetry studies, spaceborne LIDAR is not used as a reference for DSM accuracy analysis. The term “LIDAR” refers to airborne LIDAR in all references.

Usage of LIDAR is advantageous due to its high resolution and accuracy. However, public (free), global LIDAR data are not available. Thus, in this study, SRTM is used.

4.2 SRTM Registration

For the registration of the SRTM data to a single satellite image, a straightforward method is developed: *The LR-DSM data are projected to the image domain and the empty pixels are filled with quadratic interpolation.*

It is known that SRTM accuracy is unchanged when 90m SRTM data are upsampled by 3 to obtain 30m resolution, via bi-cubic interpolation

(Keeratikasikorn 2008). Therefore, we upsample the SRTM data by using bi-cubic interpolation and perform registration for the upsampled version.

Once the SRTM points are projected onto the image, an interpolation is required to fill the empty pixels. The SRTM data is regularly-sampled (1 or 3 arcseconds). On the other hand, satellite images have higher resolution and do not sample the earth surface on a regular latitude-longitude grid. $1^\circ \times 1^\circ$ SRTM patches and satellite images are never aligned. In other words, the SRTM grid does not project to another regular grid in the image domain and some SRTM points will fall outside the image, especially for narrow FOV satellite images, such as IKONOS. Thus, an interpolation scheme is required that provides acceptable accuracy, and leaves no empty pixels.

For that purpose, quadratic polynomial surface fitting is applied as follows: For each SRTM point p_s , eight neighbours of that SRTM point are projected to the image domain to define a surface patch sampled at 9 points, together with the centre point p_s (Figure 4.1). Then the projection points are taken as samples from quadratic polynomials (on image coordinates) that define a height function, a latitude function and a longitude function;

$$F_i = a_i.u^2 + b_i.v^2 + c_i.u.v + d_i.u + e_i.v + f_i \quad , \quad (4.1)$$

where

u, v = image row and column indexes,

$i = 1,2,3$ for height, latitude and longitude, respectively, and

$a_i, b_i, c_i, d_i, e_i, f_i$ = coefficients of the polynomial.

In other words, interpolation functions are defined separately for latitude, longitude and height. For numerical stability, before computing the

polynomial coefficients, the coordinates are defined in local coordinate systems by subtracting the mean of the nine points both in the image domain and in the object domain. Scale normalization is not performed to avoid unnecessary multiplications and divisions in the interpolation process.

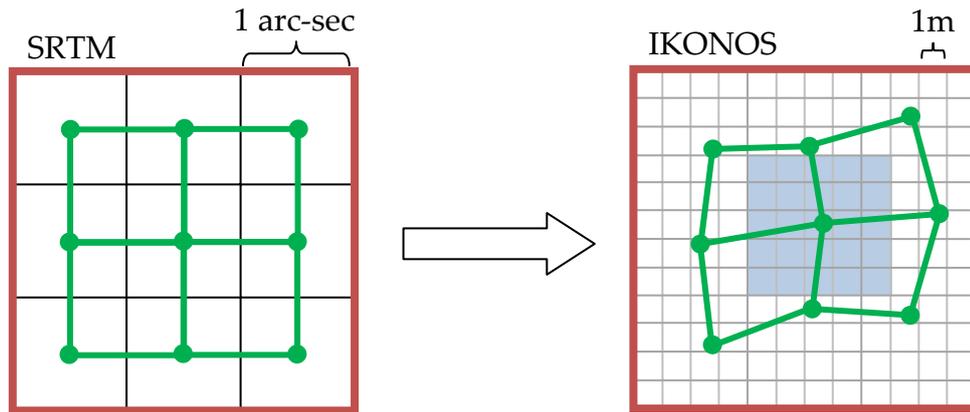


Figure 4.1. The nine-point SRTM grid and their projections on the satellite image. Interpolation is executed in the blue shaded patch

The 6 polynomial coefficients are solved by using nine equations by the singular value decomposition (SVD). The empty pixels that lie in the neighborhood of the center pixel (blue region in Figure 4.1) are filled by using the resultant polynomial. Overlapping of the neighborhoods is forced to avoid empty image pixels. For a pixel which lies in an overlap region, the interpolation value is determined using the moving average over the values that are computed for that pixel by different polynomials.

Surprisingly, this straightforward method was never published in the literature. The studies on registering satellite data to the LR-DSMs require stereo data and perform triangulation to obtain at least a sparse point cloud,

which will then be registered to the LR-DSM through correlation (*Gonçalves 2008*). In order to obtain an accurately reconstructed sparse point cloud; such approaches require very accurate projection functions, hence the GCPs. GCP-free studies cannot use such approaches since the reconstructions will be erroneous by biased RPCs. Studies on registration for a single image are absent except for already geo-registered images (*Gonçalves 2008*).

The most similar approach to the proposed one is presented in *Richards 2006*, in which a polynomial projection function (that resembles the affine projection model) is assumed from the 2D map domain (latitude, longitude) to the image domain (the height is omitted) for the LANDSAT MS sensor (50m GSD) image. The parameters of the projection function are calculated from the GCPs that are taken from the reference map or obtained from field measurements. This function is then used to generate a resampled satellite image, which is aligned to the reference map. The resampling is performed with linear or bicubic interpolation to compute the image intensities for the registered image. Obviously, the same projection function can be used in the inverse direction, for addressing each image pixel to a world point (not mentioned or described by the authors), but the mapping is 2D (latitude, longitude) to 2D (row, column). Such a mapping will result in errors for regions with steep terrain (mountainous regions), since the polynomials of latitude and longitude will not suffice for representing the effect of irregularly distributed height.

In this study, the world coordinates for each image pixel are required in both correspondence generation and reconstruction steps. The proposed registration approach is developed accordingly. Since the actual (3D-to-2D) projection functions are used during the registration process, the effects of terrain variations are handled automatically.

4.2.1 Proposed Bias Reduction Technique

As it is presented in Chapter 7, the SRTM registration with the RPCs (provided by the image vendor) has a systematic bias. The two sources of the bias are the errors in the SRTM and the errors in the RPCs.

The height errors in the SRTM are known to be larger in the steep areas and lower in the flat regions. Here, a crucial property of the SRTM is of importance: SRTM states the height value on a (lat,lon) grid without any relation to the satellite image. Thus, the geolocation error is not important for flat regions. This idea can be better understood by inspecting Figure 4.2. In this 2D example, there is a geolocation error of ΔX in the SRTM data. Similar to the SRTM case, the elevation error is higher in the steep region and smaller in the flat region. The effect of geolocation error is invisible at point C since the SRTM would have the same height value even if no geolocation error were present.

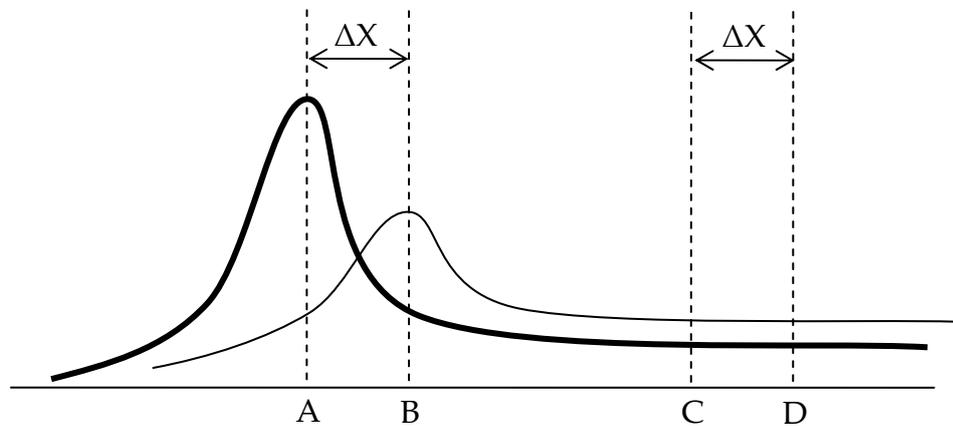


Figure 4.2. The effect of terrain shape in SRTM error: The bold curve represents the actual terrain and the light curve represents the SRTM.

The above explanation leads to the conclusion that; if one take points from the flat regions in the SRTM data, the terrain information will be correct for these points even though SRTM has overall systematic bias in geolocation. Therefore, for the points from the flat regions, the bias in the SRTM registration is caused by the RPC bias only.

The bias in RPC is mainly due to errors in satellite's position and look direction. The latitude and longitude of the satellite, as well as the look direction, are measured with some error. Although these errors affect directly the inputs of the RPCs (erroneous latitude and longitude), bias correction is usually achieved in the image domain by the help of correction terms (Δu and Δv) for both image coordinates u and v (Grodecki 2003).

Due to the above-mentioned reasons for projection bias, performing the correction in the object domain may be expected to provide better results. Still, in this study both image and object domain correction are tested.

From the previous experiments in the literature, it is known that the RPC bias is around 5 meters and not larger than 10m (for high resolution satellites, such as IKONOS) (Dial 2002a, 2002b). Additionally, coarse SRTM error figures are available for the entire SRTM coverage (Rodriguez 2005). Thus, one can determine a search region boundary for bias elimination terms.

Considering the required high computational load for satellite images, bias correction is achieved using the correspondences in the stereo satellite images by the following relatively low complexity search scheme:

- i) The reference image is registered to the SRTM data by the method described in Section 2.1.
- ii) The Kanade-Lucas-Tomasi (KLT) (Bouguet 2000) optical flow algorithm (with backward consistency) is used to determine 200

reliable stereo correspondences (tie points). The initial estimates for the KLT are obtained by taking the uncorrected SRTM registration of image 1 and projection it onto image 2 (see Figure 5.3).

iii) A quantized search is performed as follows:

Let $\mathbf{p}_1 = (u_1, v_1)$, $\mathbf{p}_2 = (u_2, v_2)$ be any correspondence between images I_1 and I_2 , and $\mathbf{P}_1 = (\text{lat}_1, \text{lon}_1, h_1)$, $\mathbf{P}_2 = (\text{lat}_2, \text{lon}_2, h_2)$ be their initial registration vectors (determined by the method described in Section 4.2). The following algorithm is proposed for determining the candidate correction vectors:

```

For each correction vector  $\Delta\mathbf{P} = (\Delta_{\text{lat}}, \theta, \theta)$ ,
    error = 0
    For each stereo correspondence pair  $(\mathbf{p}_1, \mathbf{p}_2)$ 
        Take single-image registration value  $\mathbf{P}_1$  for  $\mathbf{p}_1$ 
         $\mathbf{P}'_1 = \mathbf{P}_1 + \Delta\mathbf{P}$ 
         $\mathbf{p}'_2 = \text{Project}(\mathbf{P}'_1)$  onto  $I_2$ 
        error = error +  $\|\mathbf{p}_2 - \mathbf{p}'_2\|^2$ 
    end correspondences
end correction vectors

```

The correction vector with the minimum error score is taken as the bias correction vector's latitude component. The same procedure is applied for longitude ($\Delta\mathbf{P} = (0, \Delta_{\text{lon}}, 0)$) and height ($\Delta\mathbf{P} = (0, 0, \Delta_h)$) components of the correction term.

One may argue that the correction terms shall be computed simultaneously in a 3D search. However, this approach was tested and found to be useless. The main reason for that undesired performance is the ill conditioned nature of the inverse problem for the imaging geometry. The bias removal in a 3D search space appears to have the potential to result in a

false global minimum. Since the bias has a strong DC term for the entire image, this search is *performed only once* with a relatively smaller number of “good” tie points.

It should be noted that, this scheme determines the bias term by projection from image-1 to image-2. One can also perform bias term determination in the inverse direction and use the average of the two results; however, such a two-way method is of little use and therefore, only one-directional bias compensation is performed in this study.

Although KLT is not commonly used for correspondence generation, after proper modifications, KLT can provide impressive results yielding many sub-pixel image correspondences and relatively small computational complexity.

For the purpose of the bias reduction, one useful property of the KLT algorithm is that, it fails to find correspondences in steep regions due to occlusions and non-linear distortions. The same is true for the forests, where the observed pattern changes significantly with a change in the satellite position. It is observed that, the correspondences with the best match scores lie on bare ground in relatively flat regions; even if the flatness condition is not imposed (Figure 4.3 and Figure 4.4).

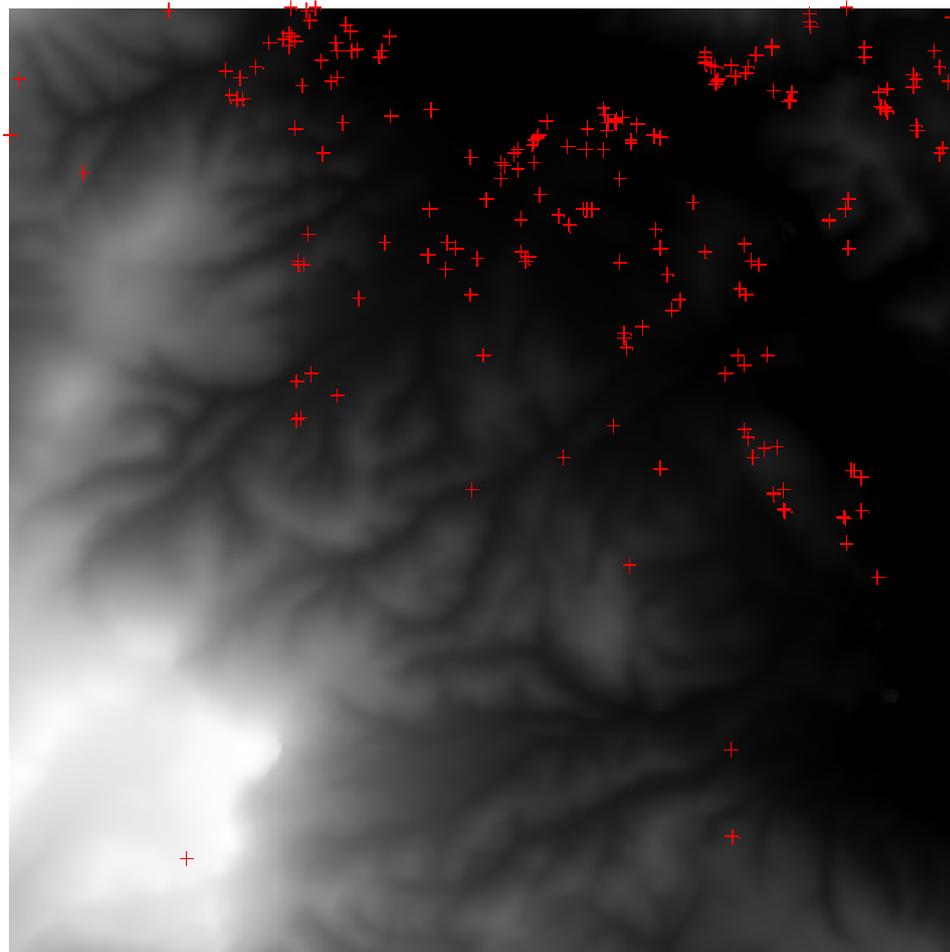


Figure 4.3. SRTM Registration result and the locations of the selected pairs for bias reduction. IKONOS image, Hobart, Australia

If the biases for both images are in the same direction, the benefits of this method are questionable for registration bias removal. At this point, it must be noted that, although LR-DSM registration is by itself useful in orthorectification and georegistration, the SRTM registration result should be evaluated with this regard, since this study aims at generating DSMs. As explained in the following sections, equal bias in the stereo pair is in fact a desirable property for the reconstruction accuracy in height. Thus, the

residual error in the SRTM registration due to equal projection biases in both images is not important.

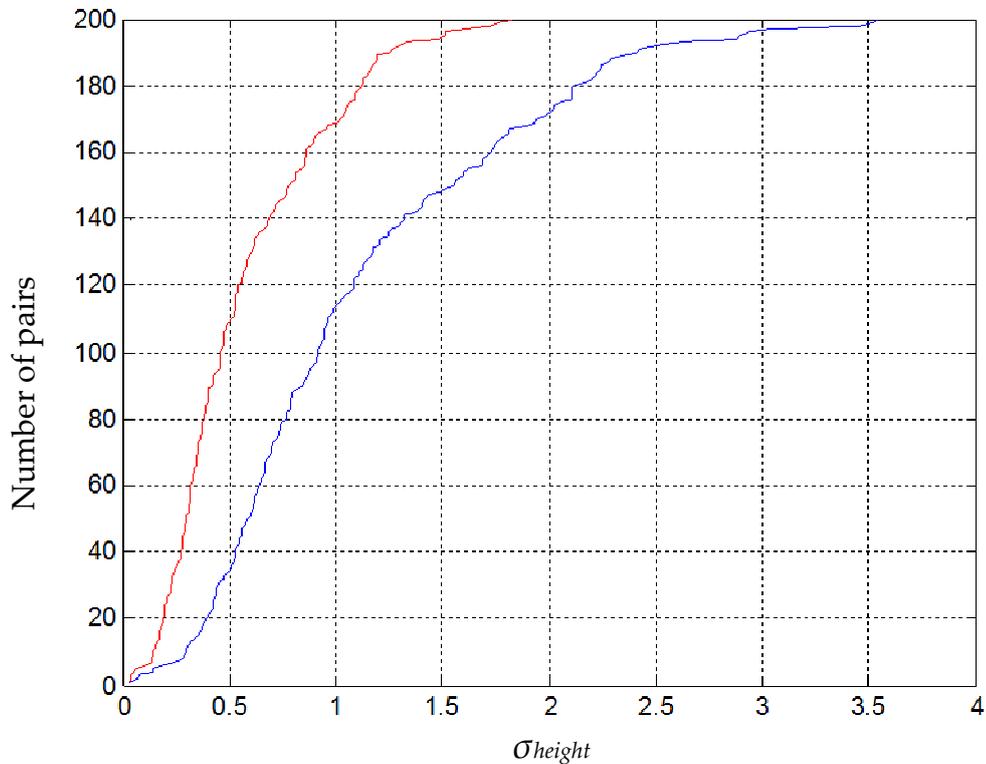


Figure 4.4. Cumulative distribution of the height variations in the vicinities of the selected pixels. The values are computed in 10m x 10m (red) and 20m x 20m (blue) neighborhoods. IKONOS image, Hobart, Australia.

In this study, SRTM registration is used *i)* for initial estimate generation in sparse correspondence search and *ii)* for obtaining the initial estimates in sparse reconstruction. Thus, bias reduction is not crucial for the final accuracy of either step. It is presented here for completeness, considering its use in georegistration and orthorectification. The basic idea in SRTM bias reduction is also used in RPC bias equalization (Section 5.3.1).

CHAPTER 5

SPARSE RECONSTRUCTION

5.1 Related Work

5.1.1 Sparse Matching

Photogrammetry Literature on Sparse Matching

The most common feature detector used in the photogrammetry literature is the Förstner detector (*Förstner 1986, Förstner 1987, Jazayeri 2008*), which is a modified version of the Harris feature detector (*Harris 1988*) that finds sub-pixel positions. The difference from the Harris detector is that, the Förstner operator checks the invertibility (or the condition number) of the structure tensor. Obviously, this is equivalent to checking the eigenvalues of the structure tensor; hence it is not different from the Harris approach. Compared to modern corner detectors in the computer vision literature, the Förstner operator is a very slow detector that generates less interest points in longer time (*Jazayeri 2008*).

Unlike general trend in the computer vision studies, most remote sensing studies prefer to use an expensive matching strategy, namely the *Iterative Least Squares* (ILS). ILS tries to find the match position together with

the affine transformation between the match regions and the radiometric correction (brightness and gain) terms (*Gruen 1985, Silveira 2008*):

$$g_2(x_2, y_2) = \alpha \cdot g_1(x_1, y_1) + \beta , \quad (5.1)$$

where α is the contrast correction term, β is the brightness correction term, $g_2(x_1, y_1)$ is the gray value at (x_1, y_1) in image 1 and $g_2(x_2, y_2)$ is the image 2 counterpart for the match point. x_2 and y_2 are given by the following affine transformation relation:

$$\begin{aligned} x_2 &= a \cdot x_1 + b \cdot y_1 + c \\ y_2 &= d \cdot x_1 + e \cdot y_1 + f \end{aligned} \quad (5.2)$$

ILS uses an iterative scheme for minimizing the gray level difference to determine the optimum affine transformation and radiometric correction parameters.

Although ILS is a very powerful technique for correspondence generation, it forces/transforms the region of interest for matching. Therefore, false correspondences can be generated if the search zone is not constrained by using the epipolar curves and height limits. Moreover, satellite images provided by some image vendors (e.g. IKONOS, ALOS) are already radiometrically corrected (*Dial 2003*).

In the stereo satellite imaging scenario, although the satellites are designed to image the region of interest with maximum overlap (intersection) for the stereo image pair, due to the non-linear image acquisition geometry and high base-to-height ratio, disparities in the stereo correspondences is large (up to a few hundred pixels). Unlike the pinhole

camera case, the imaging geometry in pushbroom imaging does not comply with the epipolar constraints. The non-linear transformation functions lead to *epipolar curves* (Kim 2000, Zhao 2008). Here, the term “epipolar” is used to imply an analogy with the epipolar lines. The correspondence search algorithms may trace the epipolar curve, provided that the projection information is accurate (Lee 2003). Epipolar resampling is rarely used. Resampling –if used- is either *i)* achieved locally, where the epipolar curve can be approximated as a straight line or *ii)* performed with simple approximations for the imaging geometry; such as parallel/orthographic projection (Morgan 2004). Even if the epipolar curve is used, due to the large image size, tracing the epipolar curve is computationally expensive because of the long search range (along the epipolar curve) due to large disparities.

Additionally, the projection information is not accurate enough to provide usable epipolar curves. Accuracy reports of modern EOSs state that with the vendor provided RPCs, the projection error is 4 to 12 pixels in both image dimensions (Dial 2002b). Usage of GCPs reduces the error down to sub-pixel level and at this point, epipolar curves can be used. For a GCP-free study, epipolar curves are less useful, since the search along the epipolar curves is required to perform the search in an image strip instead of a 1-D curve.

Consequently, obtaining initial estimates for the match points is a requirement. Various approaches are proposed for initial estimate generation. In the most widely used approach; *i)* a hierarchy based correspondence generation algorithm is applied for a small number of points (anchors), *ii)* these correspondences are triangulated *iii)* the results are interpolated to generate a coarse DSM, and *iv)* the coarse DSM is then used to generate initial estimates for the remaining correspondences (Zhang 2005).

Stereo correspondence generation is generally weak in the urban regions, especially for tall buildings (due to occlusions) and forests (due to repeated patterns) (*Zhang 2005*). Thus, to avoid false matches, it is best to keep the search region small, if possible. The hierarchical methods are usually problematic, especially in the forest regions where even the anchor correspondences are difficult to obtain.

Multiple Primitive Matching (MPM) is one of the few recent correspondence generation algorithms that are developed specifically for spaceborne and airborne images (*Zhang 2005*). This algorithm uses a union of feature points, edges and grid points in matching and applies geometric constraints (epipolarity and smoothness) on image pyramids. Feature points are obtained by thresholding the gradient vector components and matched with the correlation coefficient. The search range is determined in the object domain using the coarse DSM obtained from the higher level of the image pyramid. A square correlation window (*window 1*) is used in the reference image.

In MPM, the shape of the correlation window in the second image is not square: Its corners are determined by projecting the object domain vertices of *window 1* onto image 2, and assuming planar DSM in the neighborhood. The correlation is performed after warping the search image patch with bilinear interpolation. All primitives are matched by using this method, but for the edges, the candidate points are first eliminated by intersecting the epipolar curve with the edges in the search image. The match points and the resultant DSM are updated iteratively with probability relaxation. This is a very detailed and complicated method, which appears to be beneficial especially for the aerial images. However, it is strictly dependent on the epipolar curves and thus, requires very accurate projection functions, which can only be

obtained using GCPs. The method is used in various studies of the same research group (*Poli 2004a, Gruen 2003, Gruen 2005, Zhang 2005, Poli 2007*)

As described above, recent computer vision techniques on feature extraction and descriptor based matching have not been imported to the photogrammetry studies on spaceborne imagery. The geometric distortions are handled using geometric constraints (epipolarity) or expensive warping techniques and radiometric inequalities are handled using pre-processing steps that provide limited radiometric invariance.

Vision Literature on Sparse Matching

Correspondence generation is a grand branch of vision research. Various algorithms have been developed for different applications. Covering the entire field of study is not aimed here. However, a general overview should be presented. Comprehensive review and evaluation of the local feature detectors and the descriptors are presented in *Tuytelaars 2007* and *Gauglitz 2011*. Feature detectors used in computer vision can be divided into three main categories:

- i) Edge detectors:** Roberts (*Roberts 1963*), Prewitt (*Prewitt 1970*), Sobel (*Gonzalez 2002*), LoG, Hough Transform (*Duda 1972*), Canny (*Canny 1986*),
- ii) Corner detectors:** Harris (*Harris 1988*), SUSAN (*Smith 1997*), FAST (*Rosten 2003*), AGAST (*Mair 2010*), STAR (CenSurE) (*Agrawal 2008*), and
- iii) Blob detectors:** MSER (*Matas 2002*), SIFT (*Lowe 2004*), SURF (*Bay 2008*).

Matching strategies can be grouped in two categories:

- i) **Intensity based matching:** Template matching, KLT (*Lucas 1981*), Horn & Shunk (*Horn 1981*), Dense optical flow, and
- ii) **Feature/Descriptor based matching:** SIFT (*Lowe 2004*), SURF (*Bay 2008*), DAISY (*Tola 2010*), ORB (*Rublee 2011*), BRIEF (*Calonder 2010*), BRISK (*Leutenegger 2011*), HoG (*Dalal 2005*), Local Binary Patterns (*Ojala 1994*), Haar/Gabor features (*Daugman 1988*).

Classical computer workflow in correspondence generation starts with a feature detector in the reference image, and continues with the search for its correspondence in the search image.

Feature detectors are designed to detect interest points that pose distinctive properties and repeat in the search image with sufficient probability under several distortions (illumination change, blurring, perspective transformation, etc.).

For the matching criteria, local properties around the feature point are compared with those of the candidate points. Local properties are extracted from the intensity distribution, and are required to be both selective and robust under geometrical and radiometric distortions.

The search region is usually constrained by *i*) an assumption on the maximum translation (disparity), *ii*) utilizing the epipolar line. For speed improvements, image pyramids are frequently used (in hierarchical algorithms).

Patch size is also important in the matching process: Larger patches decrease the number of false matches (increase *precision*) while smaller patches increase the number of successful matches (*recall*). This classical dilemma can be solved partially by using a hierarchical approach, in which the effective radius of the search region and the effective patch size is gradually decreased. In this way, the probability of false matches is reduced

with large patches when the search region is large and the probability of success is gradually increased as the search region is being shrunk.

Matching is always performed with a descriptor of the feature point in the reference image. The descriptor may be the patch of intensities around the point of interest, or its image (representation) in an invariant domain. Although the invariant domain is not always a linear subspace, descriptor extraction can be better understood by the analogy to *dimension reduction* by projection onto a subspace where the vectors are invariant under one or more distortions. Dimension reduction is performed for *i)* increasing the computational efficiency of the search, *ii)* gaining robustness to various geometric or radiometric distortions and noise. It should be noted that, dimension reduction operation results in loss of information (in general) and increases the probability of false matches. However, the feature descriptors are generally designed to obtain better recall, while preserving the precision and the gain in the number of successful matches is usually larger than the loss in precision.

If matching is achieved with a descriptor that requires local operations in the description window, the search in the second image may be cumbersome when the descriptor is required to be computed for all points in the search region. This problem can be overcome if a detector with high repeatability is applied to both images and the comparisons are done only at the feature points.

The recent trend in vision research is development of fast tools that are applicable for mobile and/or real-time applications. Most algorithms are developed for the video applications, in which the change between successive video frames is limited. Thus, speed is promoted against accuracy and consequently, simpler descriptors that demand for less computational

power and smaller memory have gained popularity. Expensive but powerful detectors/descriptors such as SIFT are preferred when speed or memory limitation is not the major concern.

One of the fastest corner detectors in the literature is FAST (*Rosten 2003, Rosten 2006*), which generates the largest number of feature points in the shortest time (*Jazayeri 2008, Gauglitz 2011*). FAST uses a simple, pre-defined decision tree for deciding whether there is a corner in the point of interest. The tree branches are based on simple intensity comparisons in the neighborhood. Although FAST is fast, its repeatability is among the worst of the popular feature detectors (*Gauglitz 2011*). Nevertheless, due to the large number of the detected points in the images, there is always a FAST point (in the search image) that is very near (a few pixels away) to the actual correspondence point. Thus, if FAST is used with a descriptor which can select the nearest FAST point in the second image, the matching results can be used at least as very good initial estimates for a succeeding refinement step.

Development and evaluation of the descriptors is an active topic in computer vision. In this study, the speed and the memory requirements are of top importance, due to very large satellite images. For example, comparison of two 128-dimensional SIFT feature vectors require floating point distance measurements in a 128 dimensional space. Such high dimensionality is prohibitive when the desired number of correspondences is in the order of millions in both (stereo) images.

BRIEF descriptor is one of the fastest descriptors in the literature (*Calonder 2010*). Its memory requirement is also the lowest due to its comparison-based nature. For computing the BRIEF descriptor, a sequence of randomly selected points is generated from a Gaussian probability density

function centered at the feature point. The intensity at each point in the sequence is compared with that of its successor and a binary array is generated using the comparison results of the entire sequence. For improving its robustness against noise, smoothing by a Gaussian kernel is performed before the comparisons. It has been shown that 128 or 256 bit descriptors provide very good matching results (*Calonder 2010*). BRIEF is extremely fast in description (simple comparisons) and matching (Hamming distance: XOR + bit count).

Even the best descriptors provide limited matching performance. The recognition rate is at most 90%, even for the smallest base-to-height ratio image pairs (*Calonder 2010*). When the correspondences are generated by a detector-descriptor combination, the limited repeatability of the detector and the limited recognition rate of the descriptor result in failure for many points (no match or false matches). Thus, for the accuracy needs of DEM generation, a modified approach is required if the FAST-BRIEF pair is to be used.

5.1.2 Sparse Reconstruction

In classical computer vision, triangulation is a well established topic, due to applicability of the pinhole projection model (*Hartley 1993, Hartley 1997*). Most of the time, GCPs are not used/available and the resultant reconstruction is given as a disparity map, rather than a depth map. Obviously, the disparity map can easily be converted to a depth map, provided the relative orientation of the cameras (rotation and translation between the cameras) and the camera matrices. In such a case, the obtained depth will be in the camera coordinate system of the reference camera.

In photogrammetry, the reconstruction results cannot be in the camera coordinate system. The reconstruction is required to be in the geodetic

coordinates and for that reason, geometric calibration of the cameras is a necessity. Once the imaging model parameters are obtained, the reconstruction is possible through optimization for the sparse correspondences. For the empirical models, the derivatives are analytical, but for complicated rigorous models, that is not the case: the satellite position is computed iteratively. Therefore, for most of the time, analytical expressions are not at hand for the derivatives and numerical approximations are required.

For linear pushbroom models, the optical center and look angle is different for each image line and slight errors in these parameters result in significant errors on the Earth surface (where the lines are expected to intersect), due to the high satellite altitude. The source of error that causes non-intersecting lines might be both matching inaccuracy and errors in the used satellite parameters. Unfortunately, the match point coordinates and the satellite parameters are highly correlated. The phenomenon is explained in Figure 5.1

As it can be observed in Figure 5.1, a point may be mapped to the same pixel by different camera setup. Therefore, if there are errors in attitude (look angle), camera center position and feature matching, then the problem that considers these three errors becomes ill-posed. Thus, many rigorous remote sensing applications assume error-free feature matching (or use manual pair selection) in the image plane and try to obtain the satellite parameters that minimize the measured and calculated GCP projections, in the least squares sense. Once the satellite parameters are determined, they are accepted to be true and triangulation is performed for other image points.

Reconstruction techniques using the empirical projection models resemble the computer vision counterparts. However, the impact of Hartley's

strong theoretical contributions is mostly unobserved in the remote sensing literature.

Although reconstruction with the inverse RFMs is possible (*Di 2001*), the coplanarity constraint (intersecting the lines in the object space) should not be directly used for triangulation. In the real life situation, the projection lines do not intersect in space due to various sources of noise (spatial sampling, errors in the projection model, ephemeris error, errors in matching, limited precision, etc).

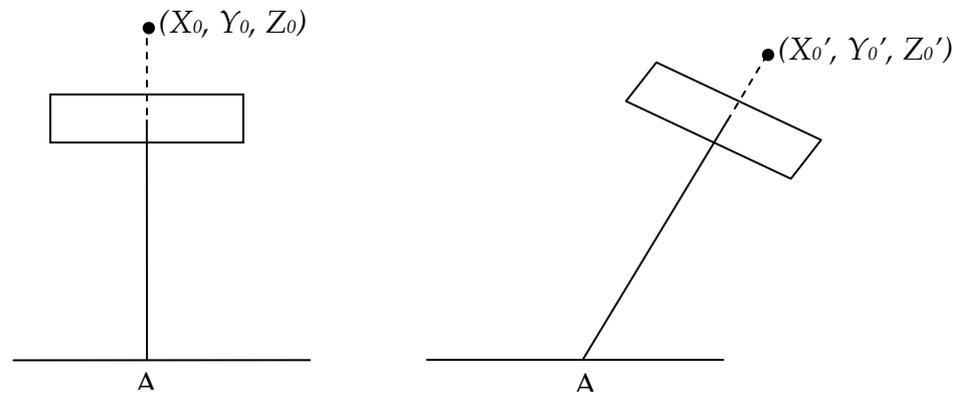


Figure 5.1. Two different cases where point A is mapped to the same image pixel.

A straightforward approach is to determine the object domain point that these two lines are closest to each other. Another slightly different approach is to use the inverse RFMs to minimize the planimetric error by optimizing the height parameter Z (*Di 2001*, *Tao 2002*). Both methods correspond minimizing a cost function that is defined in the object domain. Triangulation with object domain cost functions perform better as the B/H ratio increases. However, it provides neither affine nor projective invariance

and is strictly ruled out by Hartley for the perspective cameras (*Hartley 1997b*). In agreement with Hartley's conclusion; it has been reported that image domain cost functions provide better accuracy for satellite imagery than do object domain cost functions (*Tao 2002*). Thus, the cost function should be defined by using the image domain errors between the match points and the projections of their real world coordinate estimates.

Reconstruction with an image domain cost function is common in traditional multi-view applications of computer vision. When the projection information is perfect, this approach corresponds to refining the match point coordinates in the image plane until the two lines intersect in space (*Hartley 1993, Hartley 1995*). For the COTS cameras, the projection functions can be accurately obtained through internal and external calibration techniques which often include ground truth (e.g., checkerboards). This is correspondingly achieved by GCPs in photogrammetric applications. When GCPs are unavailable, the errors in the projection functions are often larger than the errors in the stereo correspondence locations.

Similar to the perspective case, RFMs are not one-to-one and therefore, at least two images are required for scene reconstruction. The reconstruction is achieved after adopting least squares methods by initial estimates obtained from a coarse reconstruction and/or a priori information on the average terrain height (*Tao 2002*).

The basic idea in the reconstruction is similar to the pinhole camera case: The object (world) coordinates are optimized to minimize the projection errors in the image domain. If the empirical models are used, the analytical structure of the projection functions allow closed form relations for the derivatives and simple optimization techniques are applicable for the solution (*Di 2001*). However, due to the high order cost function, the gradient

based minimization techniques may result in wrong local minima, if the iterations are started with bad initial estimates.

The closed-form solution of Hartley (*Hartley 1997b*) cannot be used in the pushbroom case due to the unavailability of epipolar geometry. Still, the epipolar curves can be used to perform matching and reconstruction simultaneously.

One alternative approach is to perform reconstruction and correspondence generation at the same time by the help of inverse functions (*Fraser 2004*). This approach fixes the geolocation (X and Y) and varies the height value (Z). For each candidate Z , the projections on both images are computed and the Z value that provides the best correlation is taken as the reconstruction. Although the original paper does not report about the convergence speed, it is expected that this method converges faster with better initial estimates. It should be noted that this algorithm does not perform actual image domain search, and thus, successful reconstruction requires projection functions with sub-pixel accuracy.

Tao proposes a similar approach, but starts the process from a feature point on the reference image (*Tao 2002*). In Tao's approach, initial estimates are used for geolocation (X, Y) and by varying the Z value. X and Y are updated iteratively by the inverse functions. The Z value that yields the highest image domain correlation is selected with the corresponding X and Y values.

5.2 Sparse Matching

In this study, a number of feature detectors and matching strategies are tested. Since the interpolation step does not include further triangulation, the sparse matcher must generate many accurate correspondences.

Sparse matching is composed of two phases: *i)* The feature detection phase and *ii)* the matching phase. Both phases are equally important for speed, accuracy and the number of generated correspondences.

If the feature detector returns many points from uniform regions, the sparse matching accuracy will be poor. On the other hand, if it is too selective, then the number of sparse correspondences will be low. Ideally, the feature detector is required to detect every distinguished point in the image in a short time.

The matcher operates on all candidate points in the search region. If the matching criterion is too strict, many correctly determined correspondences will be eliminated; and if it the criterion is too loose, many incorrect correspondences will be generated. If the descriptor is too complicated, matching might require significant memory space and computation time. On the other hand; if it is too primitive, either the selectivity or robustness (against distortions) will be poor.

The sparse matching strategy is developed with the above considerations. This phase of the study is very critical for the overall success and thus, various experiments have been conducted for obtaining the best results. The most significant approaches are presented below.

5.2.1 Interest Point Detection

In this study, for interest point detection, popular corner detectors in the vision literature are used. Most corner detectors are originally proposed for the video tracking case, in which the motion of the tracked object or the camera is small between the video frames. On the other hand, in the satellite feature matching case, the disparities may be more than a thousand pixels. However, this is the concern in matching, not in feature detection. In fact, the Förstner operator which dominates in feature detection phase of many spaceborne DSM generation studies is itself a corner detector. Thus, fast corner detectors that are able to generate many interest points could be good choices for feature detection.

Harris Corner Detector

The Harris corner detector is a very popular interest point detector that is based on the eigenvalues of the 2×2 structure tensor in the neighborhood (also known as the Harris matrix). The basic idea is that, if a corner exists in the point of interest, then the Harris matrix will have two large, positive eigenvalues (*Harris 1988*).

Since the Harris matrix is defined over a neighborhood, the neighbor pixels of an actual corner may also pass the eigenvalue check. Thus, the Harris corner detector is generally used with a local non-maxima elimination step, which forces the interest points to be separated by a minimum allowed distance.

In this study, the Harris corners are used to determine the feature points that are fed to the KLT tracker, which is known to work well on the Harris features.

Although sub-pixel extension exists for the Harris detector, it is not preferred in this study due to two reasons:

- i) The KLT results are sub-pixel. KLT finds the correct matches, without checking the cornerness of the match point. Thus, it is not required to detect the exact sub-pixel locations of the corners.
- ii) The projection functions that are used in triangulation are not accurate enough to make use of sub-pixel matches.

In this study, OpenCV implementation of the Harris corner detector is used with a minimum separation of 3 pixels.

Good Features to Track (GFTT)

The number of the feature points obtained from the Harris detector is not sufficient for healthy dense DSM generation. An alternative is proposed by Shi and Tomasi (*Shi 1994*) for the video tracking problem. This approach does not check the ratio of the eigenvalue magnitudes of the structure tensor, but only requires the smaller eigenvalue (λ_2) to be large enough. The underlying idea is that, since the digital images may take finite gray levels, the larger eigenvalue (λ_1) is bounded anyway and it is sufficient to have a large enough λ_2 to be able to track a point through the video frames.

Although Harris features are more reliable than the GFFT's for the matching task, the requirement for a large number of correspondences makes GFTT a better choice for sparse matching in the spaceborne stereo case.

In this study, the GFTT corners are used to determine the feature points that are fed to the KLT tracker, which is known to work well on the Harris features.

FAST Corner Detector

The FAST corner detector is the output of a recent study, in which machine learning is used for pruning a decision tree that decides on the corneriness, based on local intensity comparisons (*Rosten 2003, Rosten 2006*). The outcome is an algorithm which is composed of nested if-statements. The FAST algorithm is extremely fast and has the potential to generate many meaningful points. In this study, OpenCV implementation of FAST-16 is used.

The major problem of the FAST detector is that, it has insufficient repeatability for DEM generation. In other words, a FAST point detected in the reference image is not guaranteed to be detected in the second image. The repetition rate is less than 90%, even for small base-to-height ratios. Such repetition rate means that, if the match points in image 1 are searched among the FAST points of image 2, 10% of the feature points will be lost during matching; or worse; the matches for these points should have false disparity, if the matcher declares a successful match.

5.2.2 Matching

Kanade-Lucas-Tomasi Tracker (KLT)

KLT is an optical flow algorithm, which is used mostly in video applications. Although its name indicates a tracker, KLT is a feature matcher rather than a classical tracker. It finds the match of an interest point using the conventional optical flow assumption which states that the time difference between the two images (of the same scene) is infinitesimal. With this assumption, the second image can be represented in terms of the first image

via Taylor expansion with spatio-temporal derivatives, neglecting the higher order terms. The solution of the resulting equations gives the position of the match point.

KLT uses an iterative scheme to achieve the goal. Note that, KLT is not a search algorithm. It is an optical flow algorithm which solves the optical flow equation at each step. This important property makes KLT one of the fastest correspondence generation tools. However, as mentioned earlier, KLT is mostly used in object tracking rather than sparse correspondence generation in multi-view stereo. The reason is the availability of epipolar geometry, which provides sufficient speed by reducing the correspondence generation problem into a 1-D search.

Of course, even in the standard video applications (where the images are taken with at most 40ms time difference) the infinitesimal time difference assumption is not valid and the actual flow vectors may not be determined with this assumption. This problem is solved by using either image pyramids and/or initial estimates. With proper initial estimates, two or three levels of image pyramids are sufficient to obtain successful matches. In this study, OpenCV implementation of pyramidal KLT is used.

For the stereo satellite imaging case, the infinitesimal time difference assumption is far from reality. However, quite good initial estimates can be provided due to the SRTM registration and successful matches can be obtained with a few-level pyramids. This approach is applied and found successful for Harris features.

The accuracy may be improved further by applying the backward consistency check. The consistency check is performed by running the KLT from image 2 to image 1 using the forward match results as the query points. The correspondence is accepted, if the backward search finds the initial

Harris point in image 1, and is rejected otherwise. This approach guarantees the matching accuracy, but the number of successful matches decrease dramatically, since the backward consistency check eliminates many correct correspondences (see the experimental results in Section 7.5.1). Therefore, it is not suitable in dense DSM generation.

BRIEF

Binary Robust Independent Elementary Features (BRIEF) is a recently developed descriptor, which is known for its extreme computational efficiency that is obtained without significant compromise in the precision-recall figures (Calonder 2010). It is based on an ordered comparison in the neighborhood of the interest point. The points for comparison are selected randomly from a Gaussian probability density function (pdf) centered at the point of interest. The locations of the comparison points and their order is determined and saved at the beginning and all correspondences are generated with the same comparisons. Each point in the ordered list is compared with its successor and the result is put in a binary array (feature vector). The number of comparisons determines the length of the feature vector. For the matching operation, the Hamming distance between two feature vectors is used. The Hamming distance indicates the number of unequal elements. That is, it is a “element-wise XOR and count” operator. In this study, OpenCV implementation of the BRIEF descriptor is used. The selected descriptor length is 512 (64 Bytes).

The advantage of BRIEF is that, due to the Gaussian pdf that is used in the selection of the compared locations, the pixels that are close to the feature point are selected more often than the further pixels. Thus, the descriptor contains information of a large template (reduces the probability of false

matches) while it uses more pixels in the central zone, increasing the matching probability.

Template Matching

Template Matching is the simplest, as well as the strongest matching algorithm. Various matching criteria (mean squared difference, cross-correlation, correlation coefficient, etc.) can be used in template matching. The correlation coefficient is generally more useful, since it provides the ability to determine the matching threshold that is invariant to the local brightness.

Although it is a classical matching strategy, in terms of accuracy, it outperforms almost all descriptor-based matchers (which rely on the repeatability of the FAST features) in satellite stereo images of the modern satellites, when the changes in scale and rotation are minimal.

The major drawback of template matching is its long execution time. In order to guarantee matching in the urban zones, up to 40 pixels of search radius is required for high resolution satellite images. Moreover, for avoiding false matches, the template is also required to be large. The result is high computational complexity, which is undesired.

For reducing the computation time, the classical hierarchical search strategy is also applied: The two pyramids are generated from the stereo pair and in each pyramid level (upwards), the search window is halved in size, while the search window size is kept constant. The search window size is smaller than the non-hierarchical implementation.

The reason for such implementation is as follows: Large template windows are required to avoid finding the matches in incorrect locations. However, especially in the urban zones (where the occlusions are frequent), a

large template is difficult to match even if the match point is visible. In the hierarchical implementation, it is possible to select the window size that allows matching in the urban zone (in the original resolution) and keep its size constant in higher pyramid levels. Since the effective window size will be doubled in each pyramid level, the false positioned matches will be avoided unless there are repetitive patterns in the upper levels (those are very rare).

Even with the hierarchical implementation, the exhaustive nature of template matching brings significant execution time.

In this study, three-layer pyramids are used with 10x10 matching windows in each pyramid level. OpenCV implementation of template matching is used for correlation coefficient computation.

BRIF + Template Matching

The FAST+BRIF combination provides fast operation due to the limited number match candidates, but the success rate is insufficient. However, it is observed that the false matches are mostly due to the repeatability problem of FAST and the errors are mostly small disparity errors. In other words, the actual match points are quite close to the determined match points (within 4 pixels radius).

On the other hand, FAST+Template Matching provides accurate results, while the processing time is long, even for its hierarchical version.

The complementary properties of these methods bring the idea to combine them. The workflow is as follows:

- i) FAST points are detected and their BRIF descriptors are computed in both image 1 and image 2.

- ii) For each FAST point in image1, its match is searched among the candidate FAST points in image 2 with a large enough search window, using BRIEF and the Hamming distance.
- iii) Around the match result obtained in (ii), template matching based search is performed, this time with a small (4x4) search window.

This methodology provided many accurate results with good distribution in urban zones, rural areas and forests. Failure is observed for tall buildings for which both false matches and empty (no-match) zones are observed (see the experimental results in Section 7.5.1).

5.2.3 Elimination of Outliers

FAST + BRIEF + Template Matching and GFTT-KLT provides good results for sparse matching. Still, there occur many false positioned matches, which are needed to be eliminated.

The elimination must be performed with care to prevent deletion of correct matches, since insufficiency of the number of correspondences is worse than the existence of a minor set of “false matches”.

Elimination by thresholding the match score should be avoided, since such an approach should also eliminate many correctly placed correspondences, especially in the high-contrast regions. Backward consistency check should also be avoided in all cases, since it decreases the number of correspondences significantly.

The method used in study is based on *neighbor consistency* of the match points in the two images. Since most of the generated correspondences are correct, the successful match points will have the same neighbors in both images, except for the occlusion regions (tall buildings, etc). Still, due to the

many correspondences generated by FAST, even for the occlusion zones, the majority of the neighbors are preserved in the occlusion zones when the occlusion is not severe.

For the elimination of the false matches, the correspondence indices are used. Each correspondence is given an ID; and for each feature point in image 1, the ID's of the local neighbors are compared with the ID's in the neighborhood of the match point (in the second image). If more than half of the ID's are consistent, the correspondence is accepted.

This operation can be performed very fast and it has no effect in the overall computation time.

It should be noted that, the matching accuracy requirements of dense DSM generation and bias reduction/equalization steps are different: For correcting the SRTM registration and the RFMs, accurate matches are required, while for the DSM generation phase, many correspondences are needed. Thus, the elimination strategies for these phases are different. The elimination is strict for correcting the SRTM registration and the projection models and loose for DSM generation.

Outliers can be eliminated further in the sparse reconstruction phase. If the bias in the projection functions can successfully be equalized, the re-projection error of the reconstructed world point will be small for true correspondences. On the other hand, for false correspondences, the re-projection errors are larger. Therefore, a second elimination can be performed in the reconstruction phase after thresholding the re-projection errors. Obviously, the benefits of this approach are limited, since the projection functions cannot be corrected completely. Still, it is observed that many false correspondences can be eliminated using this technique.

5.3 Sparse Reconstruction

5.3.1 RPC Bias Equalization

For modern narrow FOV satellites, the RFMs have image domain bias up to ten pixels. Although reconstruction with such projection information results in errors, the error is limited thanks to the high B/H ratio in the stereo imaging scenario. Nevertheless, one should improve the RFM accuracy to obtain better reconstruction.

The bias is different in the stereo pair images. It was reported that it is sufficient to model the bias in RPCs as additive constants in both image dimensions [Grodecki 2003]:

$$\begin{aligned} S' &= S + \Delta S \\ L' &= L + \Delta L \end{aligned} \tag{5.3}$$

where S and L are the “sample” (column) the “line” (row) indexes which are obtained by the Rational Polynomial Functions. Correspondingly, S' and L' are bias corrected coordinates and ΔS and ΔL are the correction terms. Dial reports that *constant image domain bias* model is quite accurate, provided that the image cover is less than a 50km x 50km area [Dial 2002b]. The same bias correction model is used in this study.

Dial and Grodecki presents a block adjustment algorithm, which uses GCPs in the formulation. However, in this study, a GCP-free approach is aimed. Therefore, the block adjustment approach is of little use. On the other hand, of course, the success of bias removal will be limited when only the SRTM data are used (without any GCP).

Since the ultimate aim is to generate DSMs, an improvement in the RPFs is evaluated with the improvement achieved in the reconstruction accuracy. With this consideration, an alternative approach is proposed in this study. In this approach, the main aim is to reduce the reconstruction error in height. It is known that, for a stereo image pair, if the camera projection functions have the identical offset in the image domain ($\Delta L_1 = \Delta L_2$, $\Delta S_1 = \Delta S_2$), the reconstructed height (depth) will be error-free. On the other hand, the geolocation (translational) error is minimized to zero, when the offsets are identical in magnitude, but opposite in sign ($\Delta L_1 = -\Delta L_2$, $\Delta S_1 = -\Delta S_2$). These facts are demonstrated in Figure 5.2

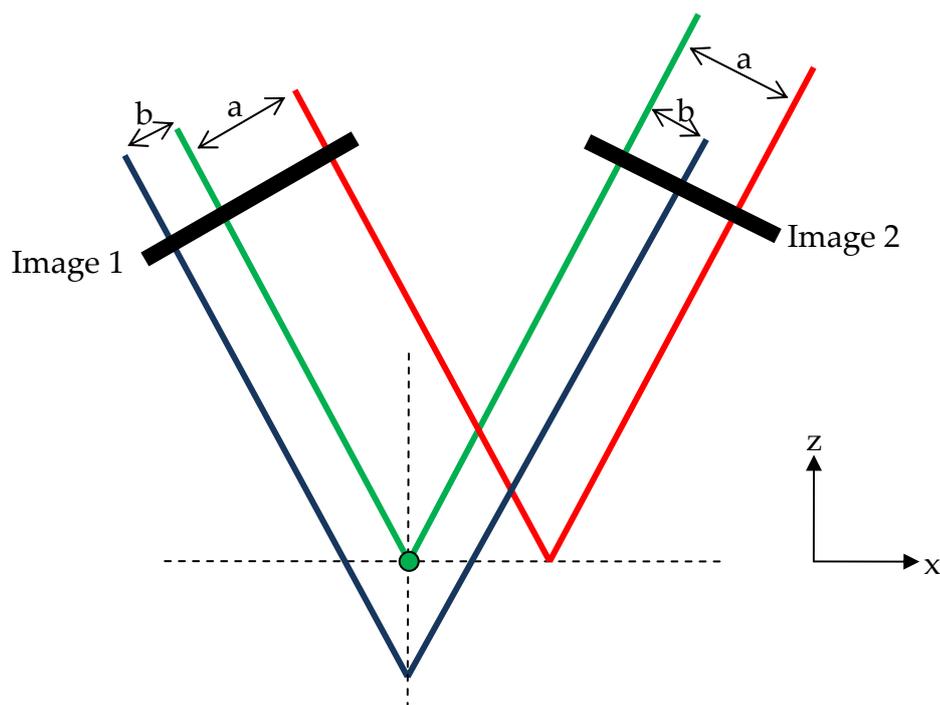


Figure 5.2. The effects of projection bias in reconstruction. The green circle is the ground point that is projected. The green lines represent the actual projections. The red lines represent the reconstruction when the projection models have the same image domain bias. The blue lines represent the reconstruction when the projection model biases are negative of each other.

In the photogrammetry literature, gradient-based reconstruction is used in general (*Di 2001, Tao 2002*). We prefer Hartley's approach during reconstruction (*Hartley 1993*), in which the sparse points are reconstructed by using the Levenberg-Marquardt optimization.

Although Levenberg-Marquardt optimization scheme is used for solving the bundle adjustment problem, it is not used in the photogrammetry literature for 3D reconstruction.

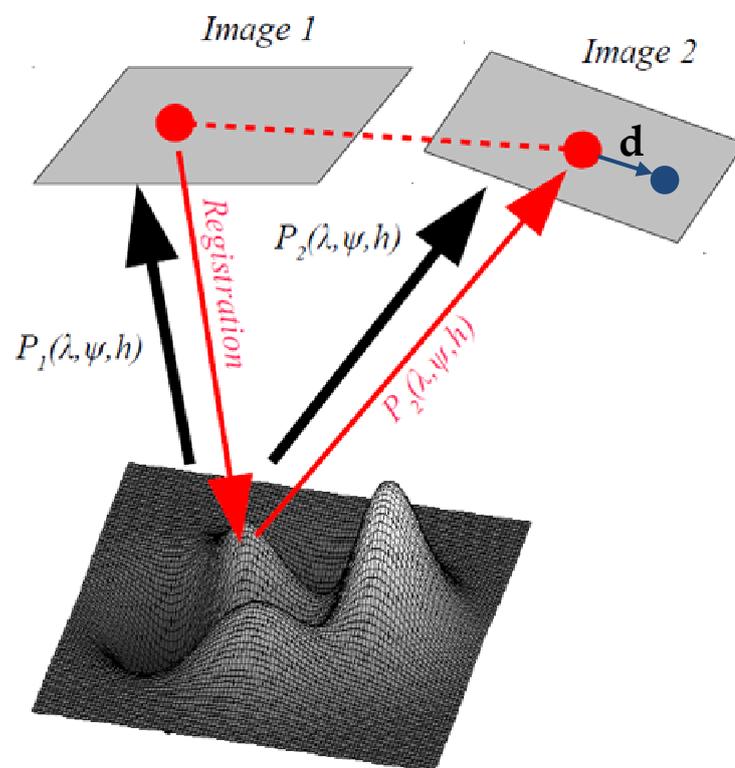


Figure 5.3. The KLT pair generation process. $P_i(\cdot)$ represents the projection function for image i . The blue arrow represents the displacement vector (d) from the initial estimate to the KLT output (blue dot).

Since the SRTM registration for image-1 is also realized with the RPCs, this algorithm is affected by the biases of both projections in a pipelined manner (Figure 5.4).

The bias terms for both projection functions are additive in the initial estimate generation process. Therefore, the difference between the initial estimate and the KLT output (blue arrow in Figure 5.3) corresponds to the difference between the bias terms. The reason will be explained by the help of Figure 5.5.

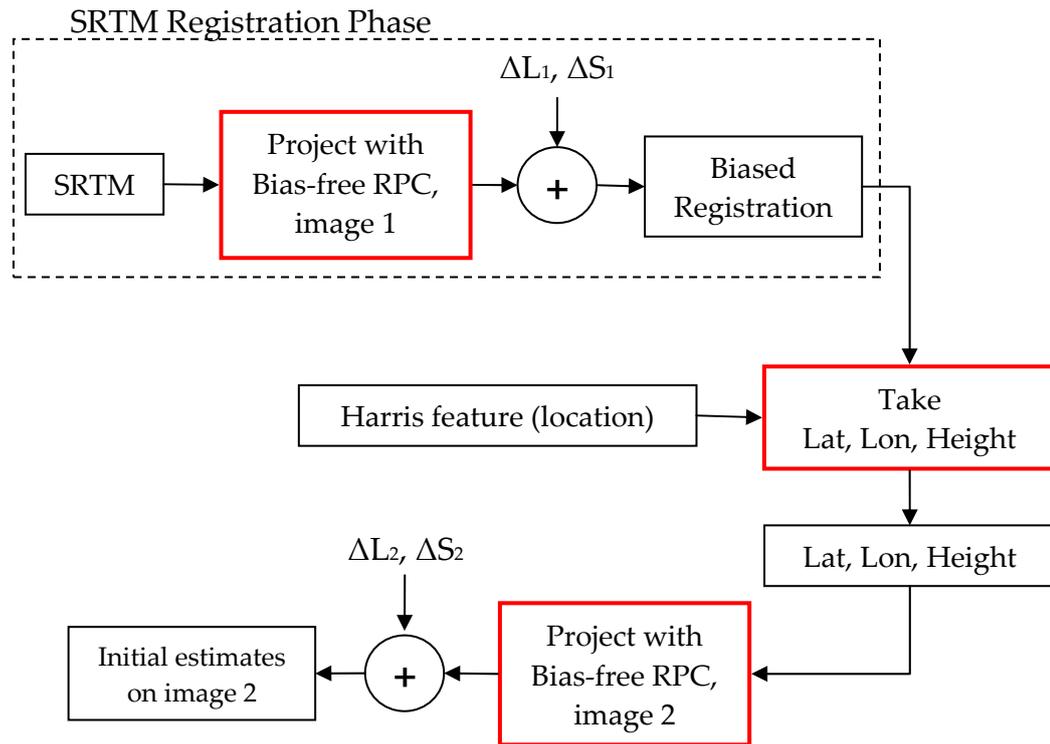


Figure 5.4. Generation of initial estimates for KLT tracker. Red boxes are operations and black boxes are data

In the scenario shown in Figure 5.5, as it is a 1D case, the direction X is along both image domain and object (world) domain coordinates. The projection functions for both images have positive bias along the X direction (image domain bias).

Point A is a Harris feature point in image 1. The geodetic coordinates for the image pixel A are taken from the SRTM registration, and projected onto image 2. This projection is fed to the KLT algorithm as an initial estimate. KLT algorithm corrects the initial estimate, ending in B as the match for A.

SRTM registration is previously performed by projecting SRTM data onto image 1 using RPCs (with positive bias ΔX_1). During SRTM registration, instead of the actual world point D, world point C is projected to A (because of bias ΔX_1). Consequently, the SRTM registration will give the coordinates of C for image point A.

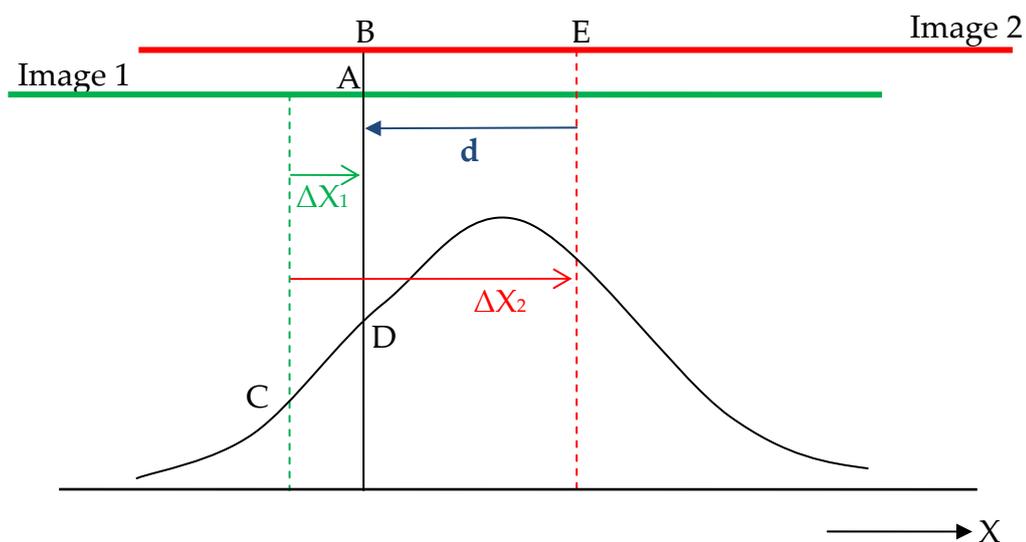


Figure 5.5. One dimensional, orthographic version of KLT initial estimate generation. Vertical black line represents the actual projection line for both images. A is the Harris corner on image 1 and B is its match to be found by KLT on image 2. The blue arrow is the displacement vector from the initial estimate to the match point in image 2

For obtaining the KLT initial estimate on image 2, C is projected onto image 2. This time, the RPC bias for image 2 (ΔX_2) is effective in the

projection and point C is projected to point E. Then, KLT takes E as the initial estimate and finds the correct match point B.

From the previous explanation, the displacement vector from the initial estimate to the correct match point is obtained to be from point E to point B (blue vector \mathbf{d} in Figure 5.5). It follows that:

$$\begin{aligned}\mathbf{d} &= -(\Delta X_2 - \Delta X_1) \\ &= \Delta X_1 - \Delta X_2\end{aligned}\tag{5.4}$$

The above explanation is valid, if *i*) the SRTM data is error-free and *ii*) the terrain that the feature point lies on is locally (within approximately 20 meters) flat. The assumptions (i) and (ii) are not valid for all points. However, the errors that these assumptions cause can be compensated for, if many pairs are used. The reasons are as follows:

- i)** For flat regions, the SRTM height error is small.
- ii)** For flat regions, the geolocation error is unimportant, since SRTM provides height information on a latitude-longitude grid and the correct geolocation should have the same height.
- iii)** If many pairs are used, the majority of these points should lie on flat or low slope terrain (except for mountainous regions). The effect of the points on high slope terrain will have less effect when the mean or median is used.

Therefore, it is reasonable to accept that the displacement vector (\mathbf{d}) from the KLT initial estimate to the KLT output is the difference of the projection functions' biases:

$$\begin{aligned}
d_L &= \Delta L_1 - \Delta L_2 \\
d_S &= \Delta S_1 - \Delta S_2 \\
\mathbf{d} &= [d_L \ d_S]^T
\end{aligned}
\tag{5.5}$$

If \mathbf{d} is the difference of the two biases, it is possible to use the idea demonstrated in Figure 5.2 to obtain better height accuracy. The bias in the two projection functions can be equalized by using the following correction terms:

$$\begin{aligned}
C_{L,1} &= -d_L/2, & C_{L,2} &= +d_L/2, \\
C_{S,1} &= -d_S/2, & C_{S,2} &= +d_S/2,
\end{aligned}
\tag{5.6}$$

where, $C_{L,1}$ is an additive term for the image 1 projection function (for line coordinates); $C_{S,1}$ is an additive term for the image 1 projection function (for sample coordinates); $C_{L,2}$ is an additive term for the image 2 projection function (for line coordinates); and $C_{S,2}$ is an additive term for the image 2 projection function (for sample coordinates)

The resulting modified projection functions are as given in (5.7):

$$\begin{aligned}
L_1 &= \frac{P_{11}(Lat, Lon, Height)}{P_{12}(Lat, Lon, Height)} + C_{L,1} \\
S_1 &= \frac{P_{13}(Lat, Lon, Height)}{P_{14}(Lat, Lon, Height)} + C_{S,1} \\
L_2 &= \frac{P_{21}(Lat, Lon, Height)}{P_{22}(Lat, Lon, Height)} + C_{L,2} \\
S_2 &= \frac{P_{23}(Lat, Lon, Height)}{P_{24}(Lat, Lon, Height)} + C_{S,2}
\end{aligned}
\tag{5.7}$$

Although this approach is useful for reconstructed height accuracy, the geolocation accuracy for the reconstruction is expected to be worse (see Figure 5.2). The biases of the two images are expected to be in the same direction for in-track stereo imaging, since the source of the RPC bias is mainly the accumulated gyroscope error, which is expected to be preserved between the two acquisitions. This might be the underlying reason for the RPC reconstruction to have more error in geolocation.

5.3.2 RPC Reconstruction

For reconstruction/triangulation, Di's method (*Di 2001*) is used with Levenberg-Marquardt algorithm. In this section, we provide the derivation of the equations and the usage of the Levenberg-Marquardt optimization for a single correspondence.

As explained in Chapter 2, the projection functions for the RPC model for image i are in the following form:

$$y_i = \frac{P_{i1}(X, Y, Z)}{P_{i2}(X, Y, Z)} \quad , \quad x_i = \frac{P_{i3}(X, Y, Z)}{P_{i4}(X, Y, Z)} \quad (5.8)$$

This formulation is in the normalized coordinates both for the image domain and the object domain. For triangulation with the unnormalized coordinates (ϕ, λ, h) and (u, v) , additional de-normalization functions are at hand (inverses of the normalization equations in Chapter 2). Thus, in general, the projection relations are:

$$\begin{aligned} u_1 &= f_{11}(\phi, \lambda, h) & , & & v_1 &= f_{12}(\phi, \lambda, h) \\ u_2 &= f_{21}(\phi, \lambda, h) & , & & v_2 &= f_{22}(\phi, \lambda, h) \end{aligned} \quad (5.9)$$

Here, the subscripts of u and v represent the image index, and the projection functions are labeled accordingly (i.e., f_{ij} stands for the j^{th} dimension of the i^{th} image, $j=1$ for u and $j=2$ for v). We define the cost function as the summation of the projection errors for the stereo correspondence:

$$E = (u_1 - u'_1)^2 + (v_1 - v'_1)^2 + (u_2 - u'_2)^2 + (v_2 - v'_2)^2 \quad (5.10)$$

In (5.10), the primed variables are the projections of the most recent world coordinate estimates (that change at every iteration) and the unprimed variables are the image domain coordinates of the stereo correspondences, whose subscripts represent the image index. It should be noted that, for a given stereo correspondence, the unprimed variables are constants and the primed variables are obtained by projecting the most recent estimates for latitude, longitude and height (ϕ', λ', h').

The Levenberg-Marquardt algorithm optimizes the following cost function:

$$E(\boldsymbol{\beta}) = \sum_{i=1}^m [\theta_i - f_i(\boldsymbol{\beta})]^2 \quad (5.11)$$

where, θ_i 's are the observations, $\boldsymbol{\beta}$ is the parameter vector that is to be determined to optimize the cost function E , and f_i is the mapping function from the parameter domain to the observation domain. In the context of sparse reconstruction; θ_i 's are the image domain correspondence coordinates ($m=4$), the vector $\boldsymbol{\beta}$ contains the geodetic coordinates (ϕ, λ, h) to be determined, and f is given by (5.9) for each image dimension (i.e., we have a different f for each observation θ_i).

The Levenberg-Marquardt algorithm for the optimization problem computes (at each iteration k) the following update vector (*Hartley 1993*):

$$\delta_k = [\mathbf{J}^T \mathbf{J} + \lambda \mathbf{diag}(\mathbf{J}^T \mathbf{J})]^{-1} \mathbf{J}^T [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}_k)] \quad (5.12)$$

where, δ_k is the update vector, \mathbf{J} is the Jacobian matrix, λ is a damping parameter, and \mathbf{y} is the vector of observations. In (5.12), bold letter \mathbf{f} means that there is a separate observation function for each observation dimension. Note that the term $[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}_k)]$ is in fact the error vector for iteration k . Therefore, in every iteration step, the Levenberg-Marquardt algorithm is converting the error vector to the parameter update vector. The iteration stops when the error norm is smaller than a threshold. In the context of triangulation, the error threshold is the acceptable total image domain projection errors in the two images. The computed update vector is used to update the parameter for the next iteration:

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + \delta_k \quad (5.13)$$

\mathbf{J} is a typical Jacobian matrix and in the case of RPC reconstruction, it is a 4x3 matrix:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial u_1}{\partial \phi} & \frac{\partial u_1}{\partial \lambda} & \frac{\partial u_1}{\partial h} \\ \frac{\partial v_1}{\partial \phi} & \frac{\partial v_1}{\partial \lambda} & \frac{\partial v_1}{\partial h} \\ \frac{\partial u_2}{\partial \phi} & \frac{\partial u_2}{\partial \lambda} & \frac{\partial u_2}{\partial h} \\ \frac{\partial v_2}{\partial \phi} & \frac{\partial v_2}{\partial \lambda} & \frac{\partial v_2}{\partial h} \end{bmatrix} \quad (5.14)$$

The relation between the image coordinates and the geodetic coordinates are as follows (explained in Chapter 2): First, the geodetic coordinates are normalized such that all coordinates of the region of interest are mapped into the $[-1, 1]$ interval. The normalization equations are:

$$X = \frac{\phi - \phi_{offset}}{\phi_{scale}} \quad , \quad Y = \frac{\lambda - \lambda_{offset}}{\lambda_{scale}} \quad , \quad Z = \frac{h - h_{offset}}{h_{scale}} \quad . \quad (5.15)$$

Normalized coordinates are then put into the projection functions which have the form

$$y = \frac{P_1(X, Y, Z)}{P_2(X, Y, Z)} \quad , \quad x = \frac{P_3(X, Y, Z)}{P_4(X, Y, Z)} \quad . \quad (5.16)$$

The projection functions produce normalized outputs. To obtain the real image domain (pixel) values, de-normalization is required. The de-normalization equations are:

$$\begin{aligned} u &= u_{scale} \times y + u_{offset} \\ v &= v_{scale} \times x + v_{offset} \end{aligned} \quad (5.17)$$

Note that, the scale and offset values of both the image domain and the object domain are different for different images (the stereo image pair have different values). These parameters are given together with the RPCs by the image providers.

The partial derivatives in \mathbf{J} are obtained through the chain rule. Here, we give the formulation only for the top-left entry to avoid redundancy (the complete formulas are presented in Appendix B):

$$\frac{\partial u}{\partial \phi} = \frac{\partial u}{\partial y} \left(\frac{\partial y}{\partial X} \frac{\partial X}{\partial \phi} + \frac{\partial y}{\partial Y} \frac{\partial Y}{\partial \phi} + \frac{\partial y}{\partial Z} \frac{\partial Z}{\partial \phi} \right) \quad (5.18)$$

The computations for $\frac{\partial u}{\partial y}$ and $\frac{\partial X}{\partial \phi}$ are straightforward. $\frac{\partial Y}{\partial \phi}$ and $\frac{\partial Z}{\partial \phi}$ are zero, since Y and Z are not dependent on ϕ . $\frac{\partial y}{\partial X}$, $\frac{\partial y}{\partial Y}$ and $\frac{\partial y}{\partial Z}$ are again determined with the simple chain rule for division:

$$\frac{\partial y}{\partial X} = \frac{\frac{\partial P_1}{\partial X} P_2 - \frac{\partial P_2}{\partial X} P_1}{P_2^2} \quad (5.19)$$

In the equation above, P_1 and P_2 are the polynomials in (5.16). The general form of the polynomials is:

$$P(X, Y, Z) = \sum_{i=0}^3 \sum_{j=0}^3 \sum_{k=0}^3 a_{ijk} X^i Y^j Z^k \quad (5.20)$$

or, in open form:

$$\begin{aligned} P(X, Y, Z) = & a_0 + a_1 X + a_2 Y + a_3 Z + a_4 X^2 + a_5 XY \\ & + a_6 XZ + a_7 Y^2 + a_8 YZ + a_9 Z^2 + a_{10} X^3 \\ & + a_{11} X^2 Y + a_{12} X^2 Z + a_{13} Y^2 X + a_{14} Y^2 Z \\ & + a_{15} Z^2 X + a_{16} Y^3 + a_{17} YZ^2 + a_{19} Z^3 \end{aligned} \quad (5.21)$$

Thus, the partial derivatives in (5.19) are analytical.

For completeness, here we give the formula for the top-left entry of the Jacobian matrix:

$$\frac{\partial u_1}{\partial \phi} = u_{1,scale} \frac{\frac{\partial P_{11}}{\partial X} P_{12} - \frac{\partial P_{12}}{\partial X} P_{11}}{P_{12}^2} \frac{1}{\phi_{1,scale}} \quad (5.22)$$

The rest of the Jacobian matrix entries are computed in the same way, with corresponding projection polynomials and scale parameters. Details are presented in Appendix B.

In our implementation, the iterations converge rapidly (in 6-8 steps) with the Levenberg-Marquardt method and good initial estimates obtained from the SRTM registration.

The value of λ is important in convergence. Large λ values make the algorithm behave like the steepest descent method, resulting in slow convergence. On the other hand, quite small λ values eliminate the effect of the regularization term, converting the algorithm into the Newton's method, which is unstable when the Hessian is ill-conditioned. In our experiments, appropriate selection of this value reduced the number of iterations from a few hundreds to a few (less than 10).

For the RFMs, we use the bias equalization terms obtained from the method described in Section 5.3.1. Note that, the formulation is unchanged for bias equalized RFM, since the correction terms are constants in the image domain (they vanish in the derivatives). The only modification is in the computation of the error vector: The bias terms are added to the projections before error vector calculation.

The initial estimates for the reconstruction are obtained from the SRTM registration for the reference image. The initial estimates can be taken from

bias-corrected registration or from the uncorrected (biased) version. SRTM bias removal is not crucial in the reconstruction phase, since the bias term can be handled during the optimization process, provided that the bias-reduced RFMs are used.

The experimental results of the all methods in this chapter could be examined in Chapter 7.

CHAPTER 6

DENSE DSM GENERATION

In this study, generating dense DSM from sparse reconstruction results (sparse points) is approached as an interpolation problem. The reconstructed points are distributed irregularly on the 2-D image. The interpolation problem is not the usual non-uniform sampling problem, since there exists the visual clue (the satellite image) that can be used as the guide in interpolation.

As explained in the introduction (Chapter 1), the photogrammetry literature on interpolation is weak. Studies that focus on the interpolation phase are non-existent. Therefore, in this chapter, we do not present a section on related work, to avoid repetition.

Dense depth generation is an active research topic in computer vision and multimedia studies, but the area of study is wide and different in nature. The dense depth generation methods utilize the epipolar lines/curves, which are not available for the GCP-free case. Therefore, the methods developed in vision-based dense depth generation research are not applicable to this study.

6.1 Meshing

The sparse points can be fed to a meshing algorithm, such as the Delaunay algorithm (*Delaunay 1937, de Berg 2008*) to obtain a triangular mesh whose vertices are the sparse points and empty pixels can be filled with linear interpolation using the values at the vertices. In the linear interpolation phase, the pixels that lie in the mesh element must be determined and special care must be taken for pixels that lie between two surface elements (on the edges of the elements).

Interpolation with meshing is edge-blind. One needs to add a mechanism to account for intensity discontinuities within a mesh element to have an edge-aware solution. Such addition is expected to bring further computational load to the method. Still, when the sparse points are sufficiently close, meshing can provide successful dense DSMs. In those regions, meshing is advantageous, since it avoids the blurring effect of content-aware algorithms. For instance, in the forest areas, many sparse points can be generated and meshing provides sharp DSMs while the color based interpolation algorithms experience problems due to color similarities. On the other hand, false matches (whose effects can be reduced by color similarity approaches) are directly reflected to a meshing-based DSM.

In this study, meshing is performed by the Delaunay algorithm, using the software named "Triangle", which is developed by the Carnegie-Mellon University and provided by U.C. Berkeley Computer Science Division (*Shewchuck 1996, Triangle web page*). The software is an efficient implementation of 2-D Delaunay triangulation algorithm.

The sparse points are interpolated by using the image domain x,y coordinates as the transversal components and the interpolated value

(latitude, longitude or height) as the z coordinate. For each mesh element, the plane passing through the 3-D coordinates of the vertices is determined and the z values of the pixels inside the element are computed using this plane. The pixels that lie in the mesh element are determined using the 2-D (image domain) line equations of the element's edges (Figure 6.1). The inequalities are determined using the vertex locations. Infinite slope lines are avoided by adding small (sub-pixel) random disturbances to the vertex coordinates and working in a local coordinate system for each mesh element.

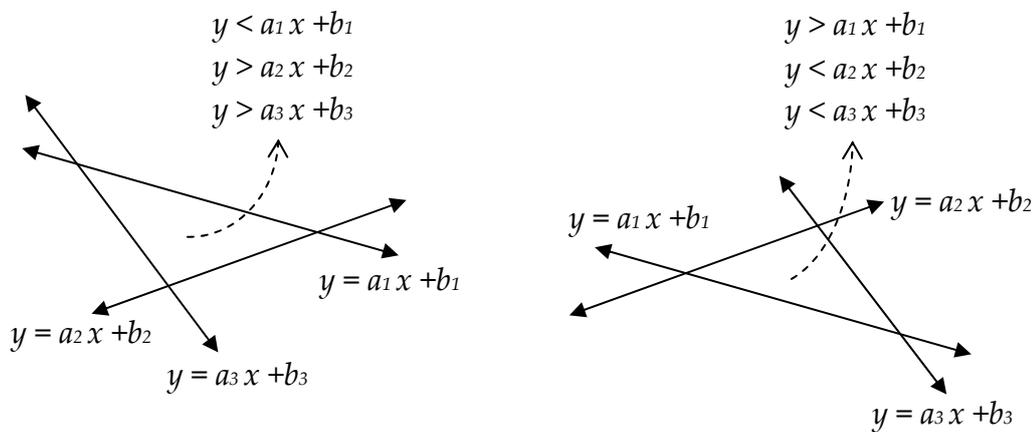


Figure 6.1. Determination of pixels inside a mesh element

6.2 Interpolation by Integral Images

The interpolation can also be achieved by a loop on empty pixels. A straightforward interpolation method is to fill the empty pixels from the available reconstruction outputs in the neighborhood using averaging. The averaging weights can again be determined using intensity difference or

geometric distance. However, determining the sparse points in the neighborhood is expensive, even if no weighting is used.

We use integral images to achieve a fast interpolation. Interpolation is performed using circular or square windows. The method is as follows:

- i) Generate an *sparse height image* whose pixels are non-zero only at the locations of sparse points.
- ii) Generate another image whose pixels are 1 at the sparse point locations and zero elsewhere (*sparse existence image*).
- iii) Generate the integral images for the two images described in (i) and (ii).
- iv) Set the range for interpolation (window edge length for the square interpolation window and radius for the circular window)
- v) At each empty pixel, sum the *sparse height image* inside the interpolation window using the corresponding integral image
- vi) At each empty pixel, sum the *sparse existence image* inside the interpolation window using the corresponding integral image
- vii) At each empty pixel, divide the result of (v) to the result of (vi) to obtain the average height value and assign it to the pixel.

Summation in a rectangular window using the integral image is well-known (Figure 6.2). Interpolation is different for circular windows (Figure 6.2). For obtaining the summation, circular window is handled row-by-row. To be able to use this integration strategy, the integral image is generated differently: In the integral image, each pixel holds the sum of the left-side pixels on the same row. In this way, the summation in each line of the circle can be computed by a single subtraction. For radius n , This method is $O(n)$

per pixel and obviously slower than the $O(1)$ the rectangular window approach.

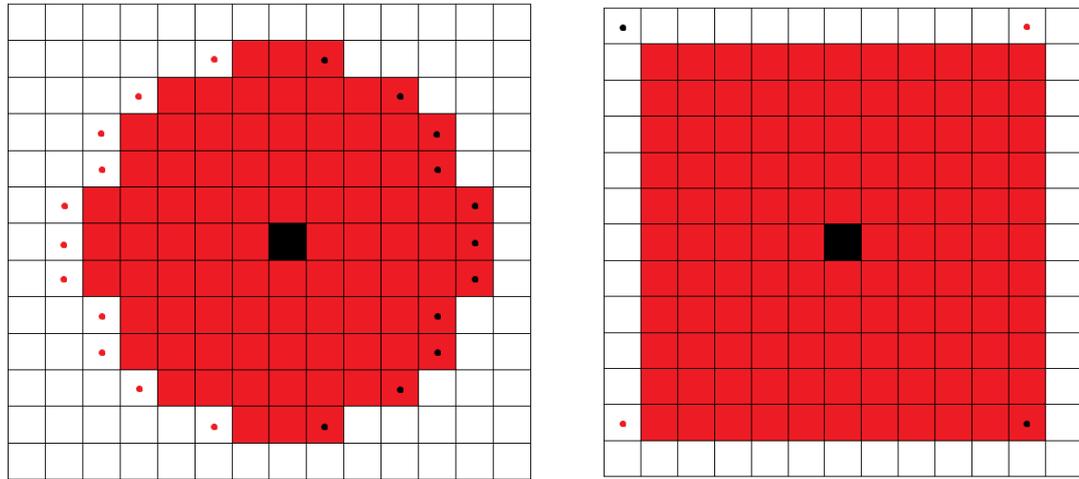


Figure 6.2 Examples for circular and rectangular windows used in integral image based segmentation. The center pixel is marked black. Window sums are computed by adding the values in black dotted pixels and subtracting the values in red dotted pixels

Usage of integral images is an edge-blind algorithm and generates blurred DEMs. However, it is extremely fast and a sharp DEM can be generated with small windows if there are many sparse points.

Interpolation by integral images can also be used as a pre-processing block for the edge-aware filtering algorithm.

6.3 Edge-Aware Filtering

Edge-aware filtering was developed by Çiğla for dense depth improvement (Çiğla 2012). In edge-aware filtering, *directed integral images* are generated with propagation weights that are determined according to color

similarity. A *directed integral image* is an integral image that is generated using 1-D summations in a pre-defined direction. Four directions are used in Çiğla's method: right, left, up and down. Right and left directed integral images are similar to the integral image used for circular windows in Section 6.2 (except the usage of propagation weights). The *right directed integral image* is generated from the left to the right, and the *left directed integral image* is generated from the right to the left. The upwards and downwards integral images are generated similarly in the vertical direction (Figure 6.3).

Using 1-D summations provide the advantage of weighting, which cannot be used in classical integral images that keep the summation of 2-D regions. The weights are computed using the following Gaussian function:

$$\omega(\Delta I) = e^{-\frac{(\Delta I)^2}{\sigma^2}} \quad (6.1)$$

where, ΔI is the color difference between the adjacent pixels and σ is an algorithm parameter which determines the sensitivity to color differences. Due to the limited number of values that ΔI can take, the computational burden of computing the exponential functions is eliminated by the help of a look-up table.

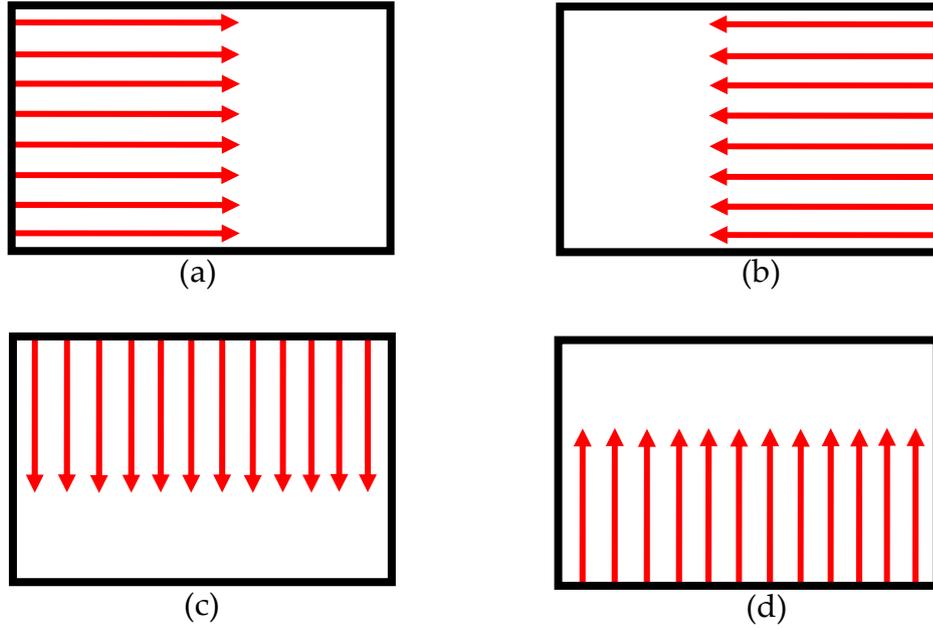


Figure 6.3. Generation of the the directed integral images in edge-aware filtering. Interpolation towards right (a), towards left (b), downwards (c) and upwards (d).

The equations for the resultant directed integral images are as follows (origin is the top-left corner):

i) The right-directed integral image:

$$S_R(x, y) = \begin{cases} D(x, y) + S_R(x - 1, y)e^{-\frac{|I(x, y) - I(x-1, y)|^2}{\sigma^2}} & , y < 1 \\ D(x, y) & , y = 1 \end{cases} \quad (6.2)$$

ii) The left-directed integral image:

$$S_L(x, y) = \begin{cases} D(x, y) + S_L(x + 1, y)e^{-\frac{[I(x,y)-I(x+1,y)]^2}{\sigma^2}} & , y < W \\ D(x, y) & , y = W \end{cases} \quad (6.3)$$

iii) The downward-directed integral image:

$$S_D(x, y) = \begin{cases} D(x, y) + S_D(x, y - 1)e^{-\frac{[I(x,y)-I(x,y-1)]^2}{\sigma^2}} & , y > 1 \\ D(x, y) & , y = 1 \end{cases} \quad (6.4)$$

iv) The upward-directed integral image:

$$S_U(x, y) = \begin{cases} D(x, y) + S_U(x, y + 1)e^{-\frac{[I(x,y)-I(x,y+1)]^2}{\sigma^2}} & , y < H \\ D(x, y) & , y = H \end{cases} \quad (6.5)$$

In the above equations; I is the image that guides the integral image generation; D is the input depth map which is of the same size with I ; H and W are the height and width for I , respectively. The propagation of the depth values is dependent on the previously visited pixels. Thus, one must keep track of the propagated weights so that proper weight normalization can be performed for each pixel to obtain the final value. This is achieved by generating a second integral image (T) from an “all ones” ($\Omega(x, y) = 1 \forall x, y$) image by applying the same procedure with the integration of depth values (i.e., in the formulas above, replace D with Ω ; and S_L, S_R, S_U, S_D with T_L, T_R, T_U, T_D).

The algorithm for edge-aware filtering uses a pipelined integral image generation process (Figure 6.4): S_L , S_R , T_L and T_R are generated from the original depth map, D and the “all ones” weight image, Ω . Before the vertical computations, the two horizontal integral images are summed to obtain the horizontal processing output ($S_H = S_L + S_R$, $T_H = T_L + T_R$). Then, S_H and T_H are used as the initial depth map and initial weight matrix in vertical processing. The vertical processing outputs are joined to obtain the final integral image, which is then divided (point-wise) by the weight integral image.

For DSM generation, the edge aware filter is used by two different approaches:

- i) The sparse height and sparse existence images are fed directly to the edge aware filtering algorithm.
- ii) The sparse height image is converted to an initial dense DSM with integral image based interpolation (Section 6.2) and the result is fed to the edge-aware filtering algorithm. The “all ones” image is used for the initial weights.

In both approaches, two problems are observed:

- i) Leakage is observed in the output. In other words, in the filtering output, the propagated heights are observed to pass beyond the edges.
- ii) Strip type artifacts are observed in low gradient regions in the guide (satellite) images, at least in the synthetic images.

One may consider that these problems are caused by large σ , however the problem cannot be solved by simply decreasing the sigma value. Another way of resolving this issue is to use a spatial attenuation constant. The attenuation constant ($0 < \alpha < 1$) is applied as a multiplier to the weight value.

This way, the effect of a sparse point is spatially bounded. The usage of attenuation constant resolves the issue for gradient regions, but the leakage problem remains.

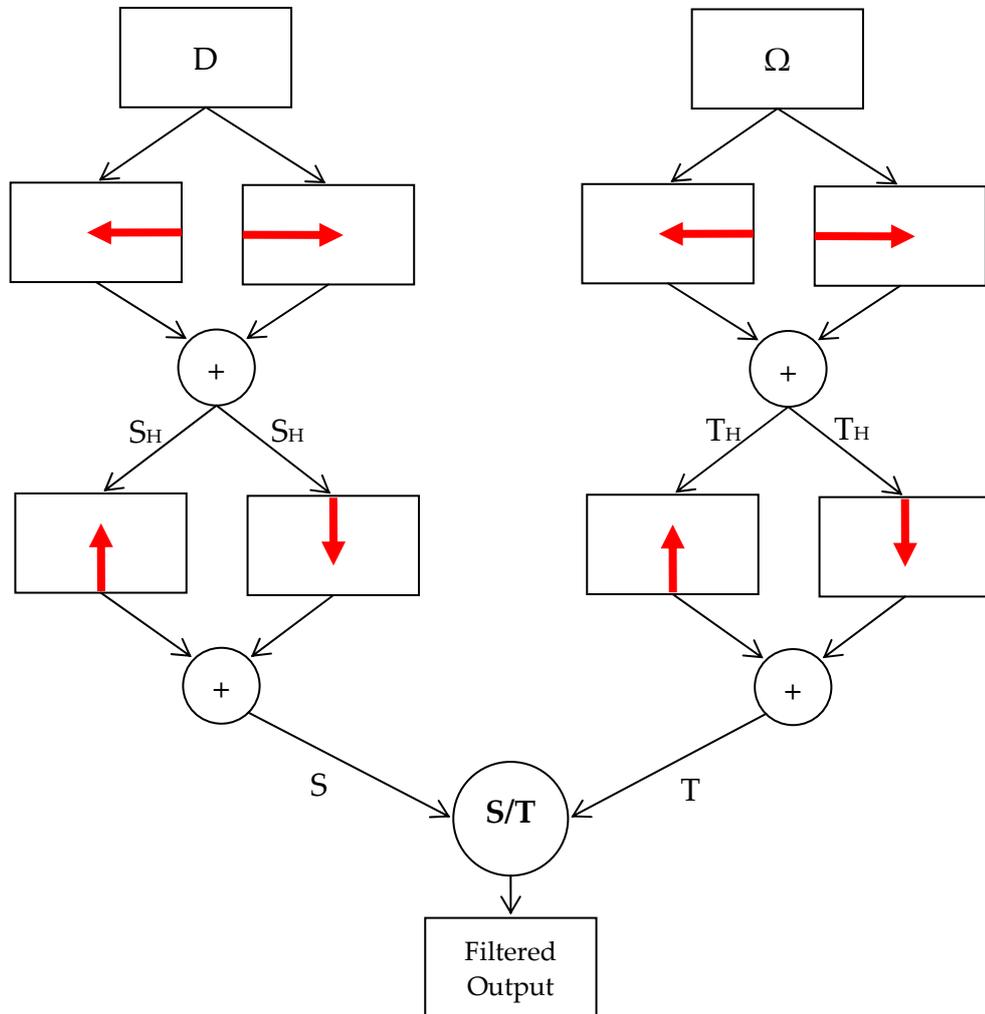


Figure 6.4. Workflow for the edge-aware filtering algorithm

The edge-aware filter can be performed with successive passes, due to its high speed. Successive passes reduce the leakage by using the values at the leakage-free pixels which constitute the majority.

The assumption in the edge-aware filtering approach is that wherever there is a disparity discontinuity, there is also color discontinuity (edge). Obviously, this assumption is not always correct; but in multimedia research, depth maps are generally used for 3D television applications. Changes in disparity are not perceived by the human eye unless there is an edge at the location of disparity change. Thus, invalidity of the above assumption is not fatal in multimedia applications.

In DSM generation, however, this is not the case. The purpose of spaceborne DSM generation is to generate an accurate DEM; the problem is not the perception problem. Performance is always measured numerically, by a ground truth.

6.4 Speeded-up Turbo Pixels (SuTP)

SuTP was originally developed for fast image segmentation (Çiğla 2010). A *turbo pixel* is a set of pixels which have both spatial and radiometric similarity. That is, the pixels in a turbo pixel constitute a set whose elements are spatially connected and have small color/intensity variance. In image segmentation, turbo pixels are used as intermediate elements which are to be joined in the later phases (graph cut, etc.) of the segmentation process. Therefore, generation of turbo pixels is classified as an over-segmentation method.

The SuTP is designed to achieve fast over-segmentation. The number of segments is determined at the beginning of the algorithm by defining a grid on the image. The grid is updated iteratively by checking the boundary pixels only. Each boundary pixel's intensity/color is compared with the average of the two segments that the boundary divides. The pixel is joined to

the segment that is more similar, and the boundary is updated (or kept unchanged) accordingly. The algorithm runs iteratively and the number of iterations determines the final output: using a few iterations preserve the organized structure of the grid and the size of the segments, and using more iterations fits the segments to the edges better.

In this study, SuTP segmentation is used in dense DSM generation as follows:

- i) For each SuTP segment, enclosed sparse reconstruction points are determined.
- ii) The average height of the enclosed sparse points is assigned to every pixel of the segment.
- iii) The empty segments are filled using the heights of the neighbor pixels/segments which are more similar in color/intensity.

The SuTP output is a buffer which has the same size with the segmented image. Its pixels keep the indexes of the segments that they belong to.

The usage of SuTP in dense DSM generation resembles the integral image approach; but this time, the main loop is run on the segment list rather than the empty pixels. Again the *sparse height image* and the *sparse existence image* are generated. Then, for each SuTP element (segment), the average of the sparse points within the segment is assigned to all segment pixels.

This method is very suitable for dense DSM generation from sparse reconstruction outputs, since *i)* the grid-based nature prevents very wide segments automatically and *ii)* the segmentation algorithm preserves edges. However; problems occur if sparse points are few. In that case, there will be many segments which do not contain any sparse point.

CHAPTER 7

EXPERIMENTAL RESULTS

7.1 Stereo Datasets Used in the Thesis

Stereo satellite data are not public, and freely available datasets for research are limited. Some of the freely available stereo image pairs are not accompanied by GCPs, thus accuracy assessment is not possible for the results on these images. The available datasets that were used in this study are the following:

- i) IKONOS triple images for Hobart, Australia,
- ii) WorldView-1 (WV-1) stereo, Spain: La Mola, Terrassa, Vacarisses,
- iii) CARTOSAT-1 (CS-1) stereo, Spain: Spain: La Mola, Terrassa, Vacarisses,
- iv) IKONOS stereo, San Diego, California, USA
- v) IKONOS stereo, Eskişehir, Turkey,
- vi) ALOS triple, Norikura, Japan (Mountainous), and
- vii) ALOS triple, Hiroshima, Japan.

For all regions, SRTM data are available. The majority of the experiments are performed on IKONOS Hobart data, which provide regional diversity (forest, urban, rural, and industrial areas, lakes, sea and cloud cover), significant height range (0-1200m) and many GCPs with a good

distribution in geolocation and height. This dataset is provided by the ISPRS and is widely used in photogrammetry studies.



IKONOS Hobart

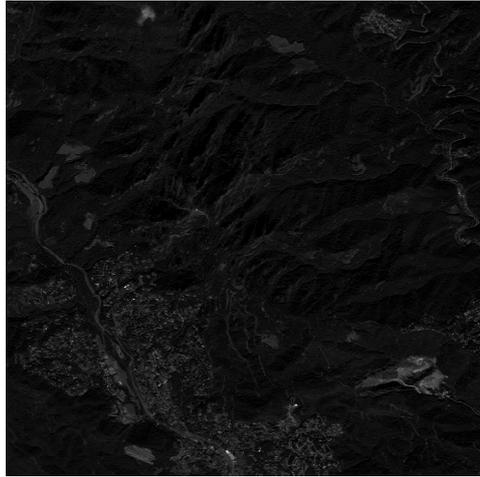


IKONOS San Diego

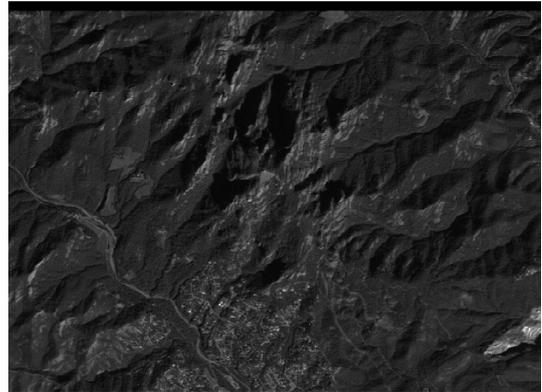


IKONOS Eskişehir

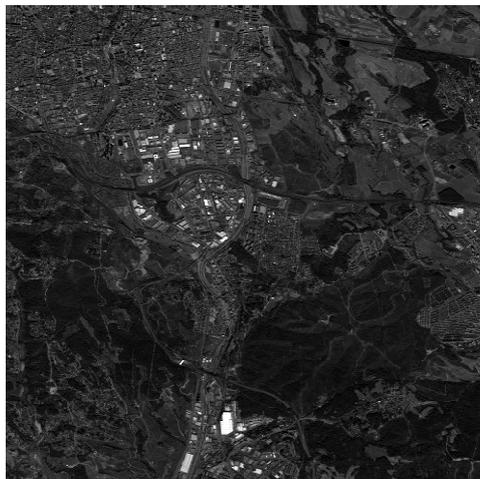
Figure 7.1. IKONOS images (enhanced for better display)



WorldView-1 La Mola



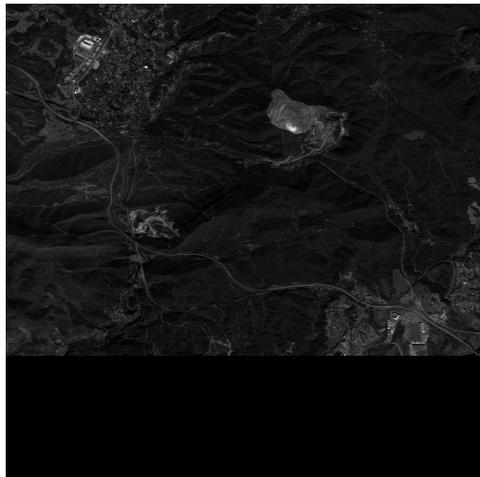
CARTOSAT-1 La Mola



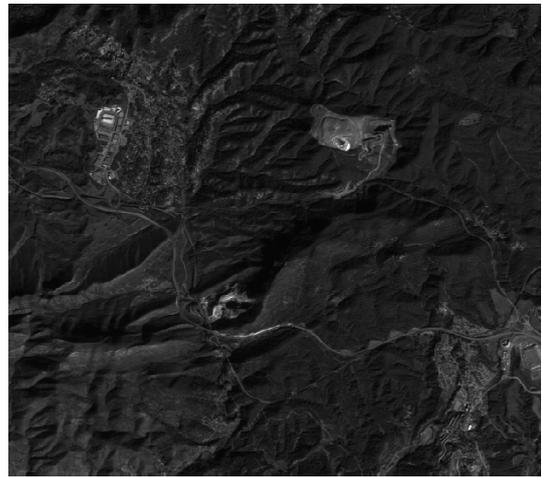
WorldView -1 Terrassa



CARTOSAT-1 Terrassa



WorldView -1 Vacarisses



CARTOSAT-1 Vacarisses

Figure 7.2. WorldView-1 and CARTOSAT-1 images (enhanced for better display)

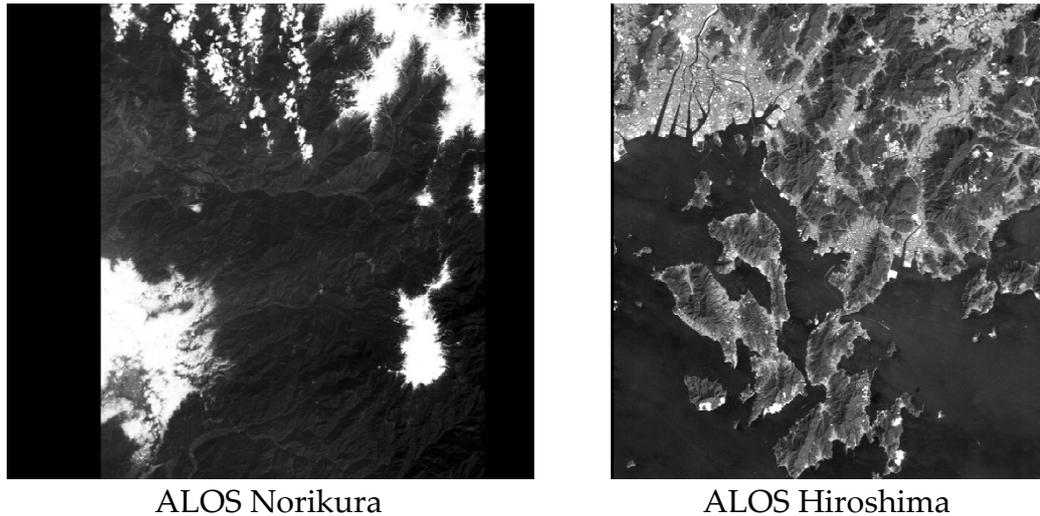


Figure 7.3. ALOS images (enhanced for better display)

The used datasets vary in contrast, dynamic range, base-to-height ratio, the existence of nadir images, the existence of color, GSD, radiometric consistency, projection model accuracy, and land cover. Thus, the dataset is a good set for determining the potential and limits of the proposed approaches.

The IKONOS images (1m GSD) have nadir views, good contrast and radiometric consistency between the images. In the Eskişehir dataset, there are many similar rooftops and many occlusion zones, which may confuse the sparse correspondence generation algorithms.

CARTOSAT-1 images (2.5m GSD) have small dynamic range (80 gray levels), poor contrast and significant noise. The accuracy of the provided RPCs is also poor. The nadir images do not exist, and there is a significant geometric non-linear distortion between the stereo images (especially in the x direction).

The ALOS images (2.5m GSD) have nadir views. In the Norikura dataset, there exists a significant radiometric difference between the two

images and between the adjacent columns of the same image (Figure 7.4). The contrast is poor and the SNR is low. On the other hand, the Hiroshima images have relatively better contrast and less radiometric difference for the stereo pair, but compression artifacts are observed in the low-contrast regions (Figure 7.5).

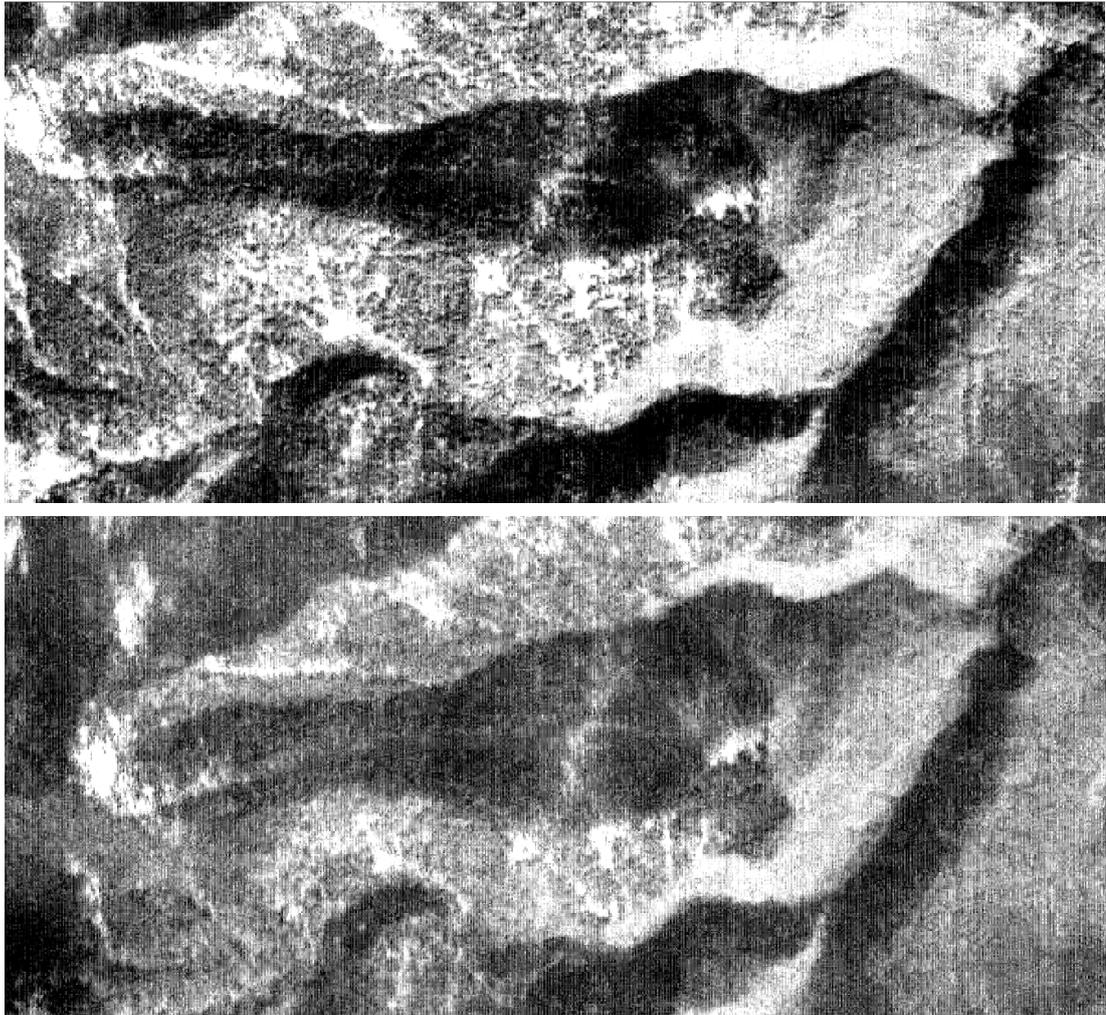


Figure 7.4. ALOS Norikura images after Wallis filtering with the same parameters. The radiometric inequality between the image columns are visible. The contrast difference between the stereo pair is caused by radiometric inequalities.



Figure 7.5. Compression artifacts observed in the ALOS Hiroshima stereo pair. The artifacts are visible in the low contrast regions (the river and shadows).

WorldView-1 images have nadir views. The contrast is good in the urban regions, but poor in the forests and mountainous areas. There is a significant geometric disturbance and scale change between the stereo pair. That is, the image images have different ground sample distances. A radiometric inconsistency is also observed (Figure 7.6). The stereo image pair

has severe occlusions and geometric differences (Figure 7.7). These problems are prohibitive in correspondence generation, even when epipolar curves are available.

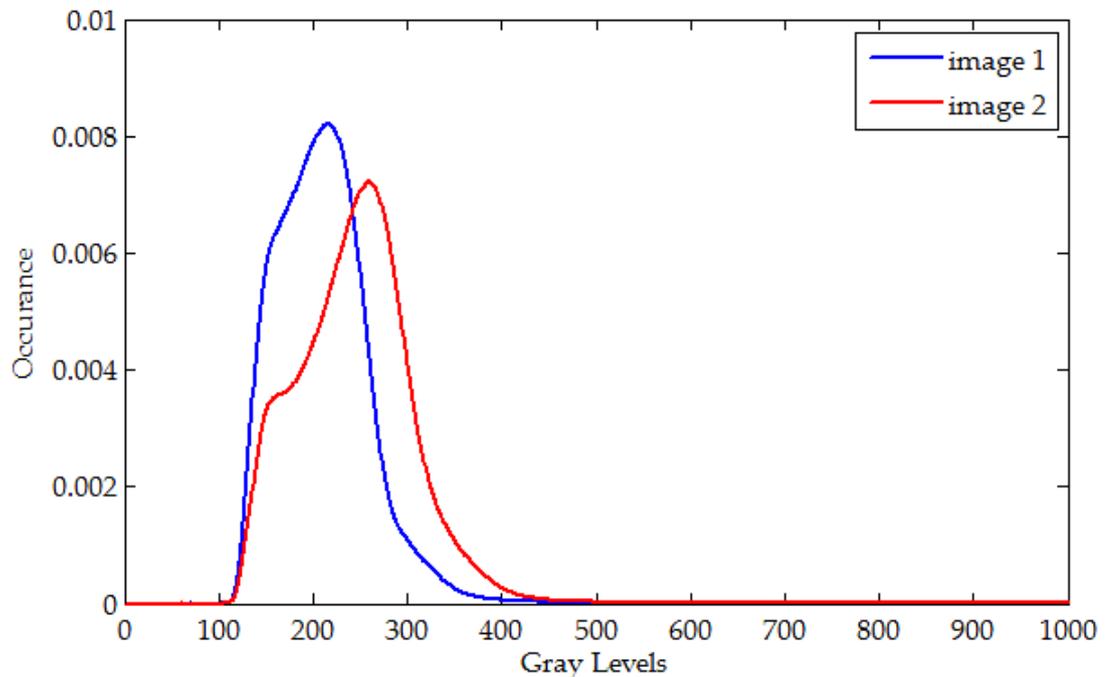


Figure 7.6. Normalized histograms for the overlapping region of WV-1 La Mola images. Histograms are normalized to overcome the unequal number of pixels in the overlap region.

7.2 Ground Control Data

Two types of ground control were available:

- i) IKONOS Hobart sequence : 122 GCPs
- ii) Spain datasets: image domain bias correction parameters, obtained from undisclosed GCPs.

GCPs were not available for other datasets.



Figure 7.7. Two patches from the WorldView-1 La Mola stereo pair. The original 11-bit images are converted to 8 bits by simple division. The digital zoom levels of the images are the same.

7.3 Accuracy of the Provided RPCs

The image domain bias in the projection functions is classically measured by comparing manually marked GCPs with their projections in the images. GCPs are provided with the IKONOS Hobart images. The RPC bias of the CARTOSAT and WV-1 images are provided along with the images.

For the CARTOSAT case, the RPC correction information is provided as an affine transformation, rather than simple bias term. Although the transformation is bias-dominant, affine effects are significant in the bottom and right parts of the images. These error figures are presented in Table 7.1. However, the RPC errors for the ALOS images cannot be reported here, since GCPs or RPC accuracy information are not provided for the ALOS images.

Table 7.1. The image domain error figures for the projection functions provided with the datasets

		Image 1, X (pixels)	Image 1, Y (pixels)	Image 2, X (pixels)	Image 2, Y (pixels)
IKONOS Hobart	μ	-5.59	-4.10	-3.46	-1.65
	σ	0.798	0.748	0.879	0.794
WV-1 Vacarisses	μ	0.45	-7.30	-9.56 -9.99*	0.19 0.11*
WV-1 Lamola	μ	0.45	-7.30	-9.56	0.19
WV-1 Terrassa	μ	0.64	-7.15	-9.99	0.11
CS-1 Vacarisses	μ	-55.17	-29.31	-97.05	-34.85
CS-1 Lamola	μ	-54.76	-28.05	-97.09	-33.03
CS-1 Terrassa	μ	-54.85	-27.31	-97.32	-32.11

* The second image in the Vacarisses data is divided into two image tiles and separate RFMs and correction terms are provided for each tile.

7.4 SRTM Registration

7.4.1 Filling Gaps in the SRTM Data

In the SRTM data, up to 5km x 5km gaps were observed, especially in the mountainous areas. Obviously, such large gaps cannot be filled without using additional data for the region of interest. Still, smaller gaps exist and can be filled using the local SRTM data.

In this study, the gaps in the SRTM were filled with an iterative technique, in which the empty pixels were filled gradually starting from the non-empty neighbors along the rows and the columns. Filling was performed using the slope between the nearest non-empty pixels in the row or the column. In each iteration step, only a limited number of pixels were filled in the row or column of interest to allow the other near-by gradients to be effective in the next iteration.

This algorithm could successfully fill the gaps up to a few hundred meters and provide intermediate results for larger gaps. Gaps larger than a few kilometers could be filled successfully. The result for the Mount Norikura region is presented in Figure 7.8. Note that, SRTM registration is only an intermediate step and the height accuracy in the filled regions is not decisive in the final DSM accuracy, provided that the filled values are accurate enough to provide good initial estimates in the matching and the reconstruction phases. In fact, in the regions where the experiments were conducted, the SRTM data do not have such large gaps.

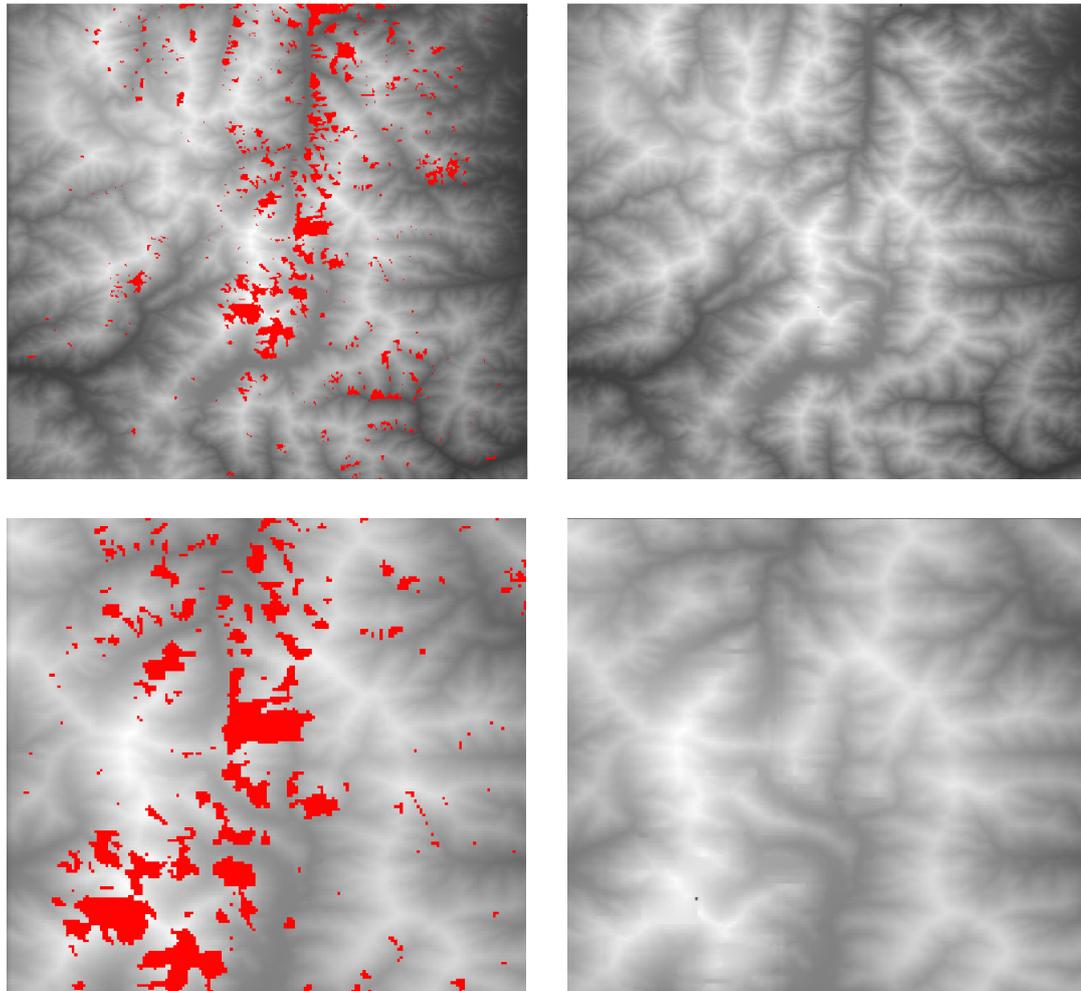


Figure 7.8. Gaps (shown red) in the SRTM and the filling result, Mount Norikura, Japan. The second row is the zoomed version of the first row. The large gap near the center of the bottom-left image is 2.7 km wide

7.4.2 Registration

SRTM registration was performed for all datasets as the first step, using biased RPCs that are provided with the images. The used SRTM patches and the coverage of the satellite images are shown in Figure 7.9 for IKONOS Hobart and ALOS Norikura datasets.

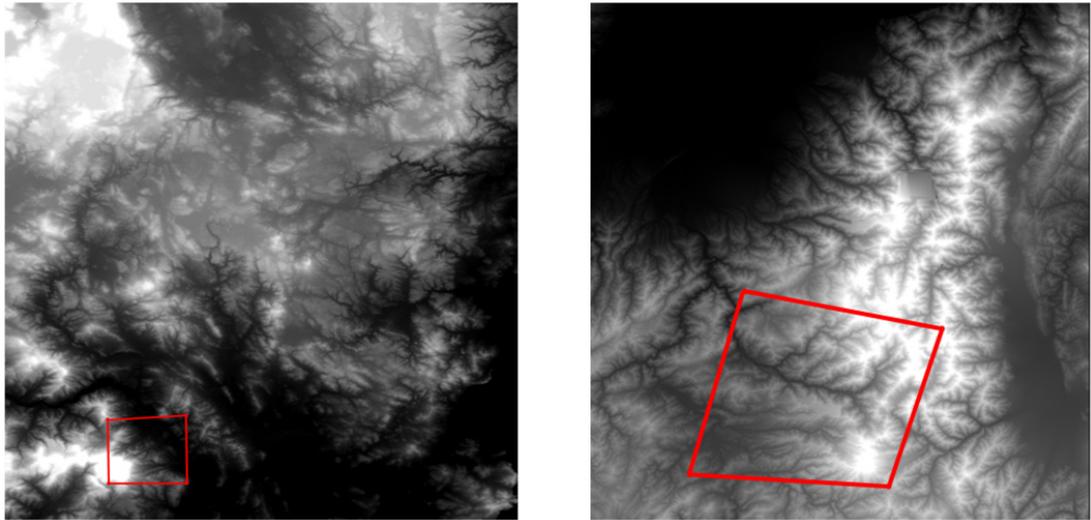


Figure 7.9. SRTM patches and image coverages for IKONOS Hobart (left), and ALOS Norikura (right) images

The accuracy of the SRTM registration is important for *i*) understanding its potential in initial estimate generation (feature matching and sparse reconstruction) and *ii*) understanding the potential of RPC bias equalization, which uses the registered SRTM. The accuracy of the registration was measured again at the 122 GCPs of Hobart region, with and without bias reduction (see Section 4.2.1). The results are presented in Table 7.2 and Table 7.3

Table 7.2. Error figures for SRTM registration without bias reduction for IKONOS reference image, Hobart

$\text{Value}_{\text{GCP}} - \text{Value}_{\text{registration}}$	Mean error	σ_{error}	RMS error
Latitude error	$2.13 \times 10^{-5} \text{ }^\circ$	$1.54 \times 10^{-5} \text{ }^\circ$	$2.624 \times 10^{-5} \text{ }^\circ$
Longitude	$-5.56 \times 10^{-5} \text{ }^\circ$	$1.46 \times 10^{-5} \text{ }^\circ$	$5.751 \times 10^{-5} \text{ }^\circ$
Height	-5.3 m	4.28 m	6.80 m
Planimetric error	5.28 m	1.64 m	5.28 m
Along UTM X	-4.54 m	1.15 m	4.68 m
Along UTM Y	2.43 m	1.65 m	2.93 m

Table 7.3. The error figures after bias removal along each geodetic axis, for IKONOS reference image, Hobart. The corresponding correction terms are; latitude: $2.3 \times 10^{-6} \text{ }^\circ$, longitude: $-3.4 \times 10^{-5} \text{ }^\circ$, height: -5.68 m

Value_{GCP} – Value_{registration}	Mean error	σ_{error}	RMS error
Latitude error	$1.90 \times 10^{-5} \text{ }^\circ$	$1.54 \times 10^{-5} \text{ }^\circ$	$2.44 \times 10^{-5} \text{ }^\circ$
Longitude	$-2.16 \times 10^{-5} \text{ }^\circ$	$1.46 \times 10^{-5} \text{ }^\circ$	$2.61 \times 10^{-5} \text{ }^\circ$
Height	0.38 m	4.28 m	4.28 m
Planimetric error	3.01 m	1.66 m	3.44 m
Along UTM X	-1.76 m	1.15 m	2.10 m
Along UTM Y	2.17 m	1.65 m	2.72 m

As seen in the results, the bias reduction provided 5m improvement in mean height error and 2m improvement in the in the mean planimetric error for the Hobart region. Since the SRTM is a low resolution DSM, it provides better initial estimates than do the DTMs for correspondence generation and reconstruction.

The registration errors were low for IKONOS and ALOS images, but for CARTOSAT and WorldView-1 images (which are obtained from the same region) the errors were very large. For the WorlView-1 images, the registration errors were up to 80 pixels (40m), and for the CARTOSAT images, they were even larger. Such large errors are caused partly by the errors in the projection functions and partly by the geolocation errors in the SRTM. Especially for the CASTOSAT-1 images, the registration is almost useless due to the projection errors that correspond to more than 250m error on the ground. For the WorldWiew-1 images, ISPRS reports at most 5m error, thus the only source of large registration error is the SRTM geolocation error. Although the reported mean absolute SRTM error for the Eurasia region (8.8m) is not as high, SRTM is known to have larger errors in the

mountainous regions. Additionally, the SRTM has 90m GSD in the Eurasia zone; thus in fact, 40m geolocation error is sub-pixel for the SRTM.

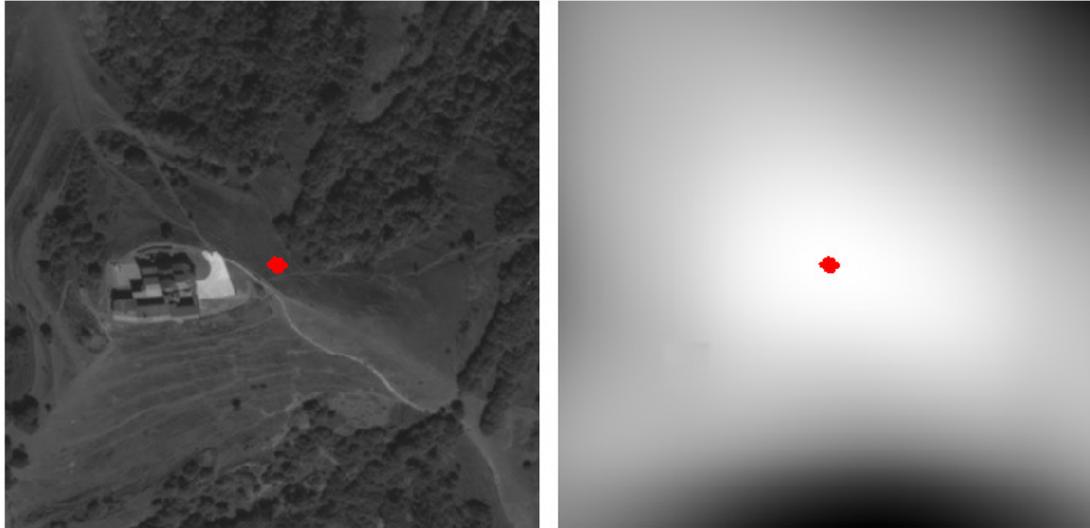


Figure 7.10. The registration error for the WorldView-1 La Mola image with the corresponding registration result on the right. The red dot is the peak location in the registered SRTM. The correct location is the north (upper) side of the building. The displacement between the correct location and the registration peak is 80 pixels (40m).

The SRTM registration results for tested satellite images are presented in Figure 7.11 -Figure 7.14. The RPC data account for the black regions in the images, thus the registration outputs cover the entire image areas, including the black zones in the Norikura and Eskişehir images.

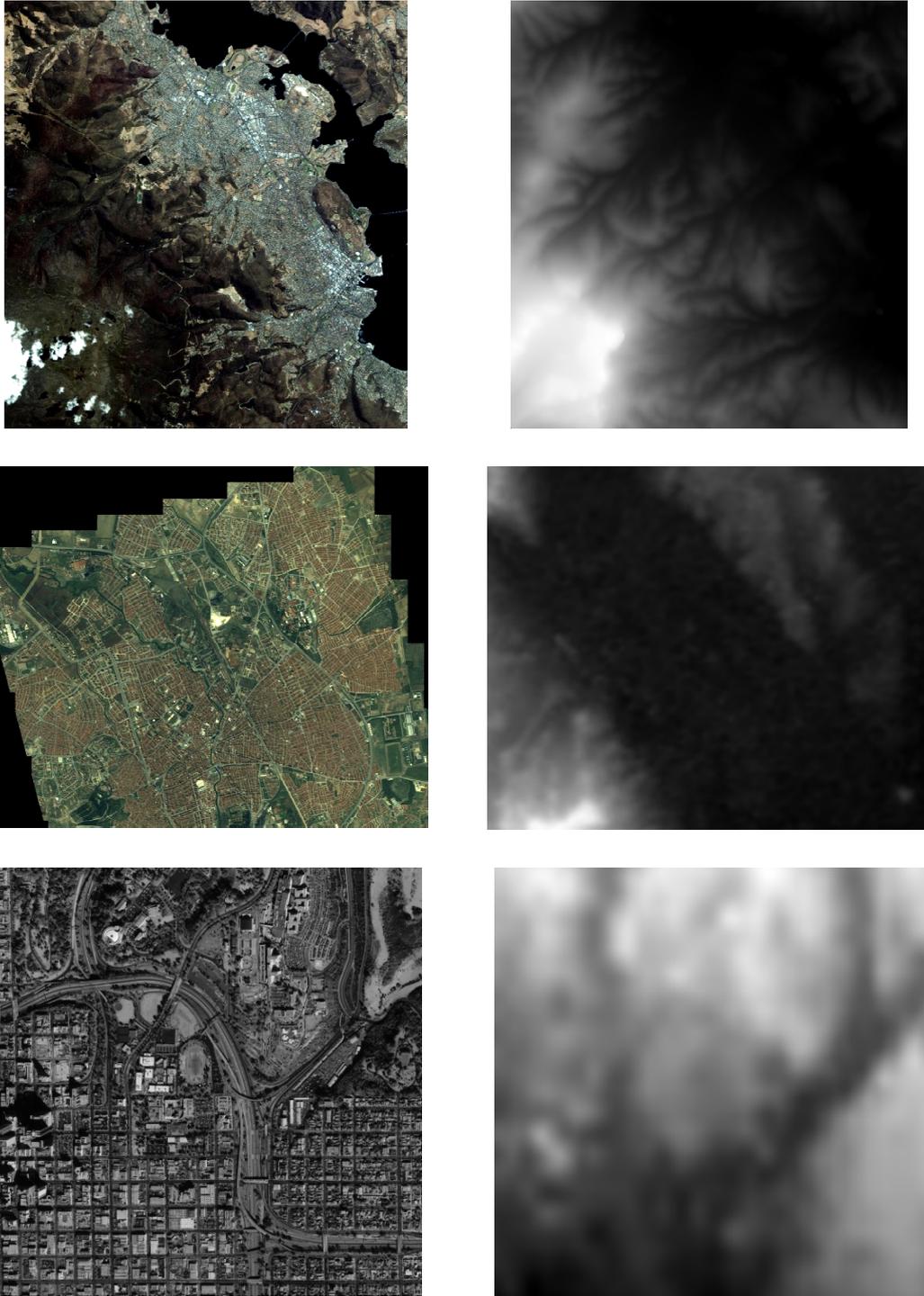


Figure 7.11. SRTM registration results for IKONOS Hobart (first row), Eskişehir (second row) and San Diego (third row) reference images (enhanced for better display)

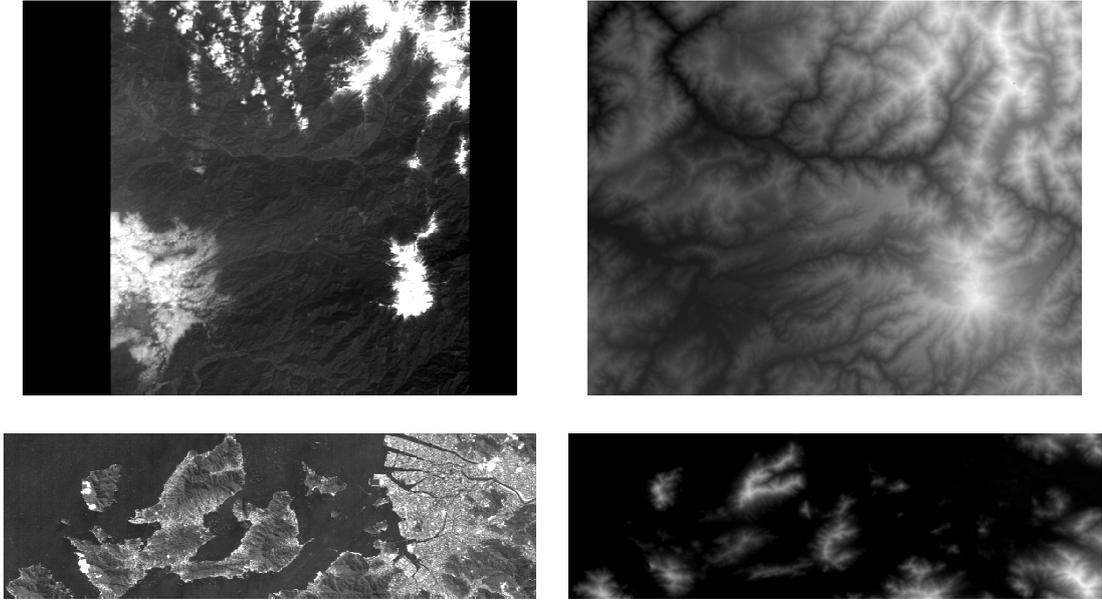


Figure 7.12. SRTM registration results for ALOS Norikura (first row) and ALOS Hiroshima (second row) reference images (enhanced for better display)

Bias reduction was successful in the Hobart region, due to the dependable (high accuracy) correspondences. However, its success is not guaranteed for other regions (e.g., WorldView-1 La Mola), when accurate correspondences cannot be guaranteed. Still, in such cases, a few manually entered tie points will be sufficient for bias removal, provided that the tie points have good height and geolocation distribution. Obviously, in this case, the workflow will be semi-automatic.

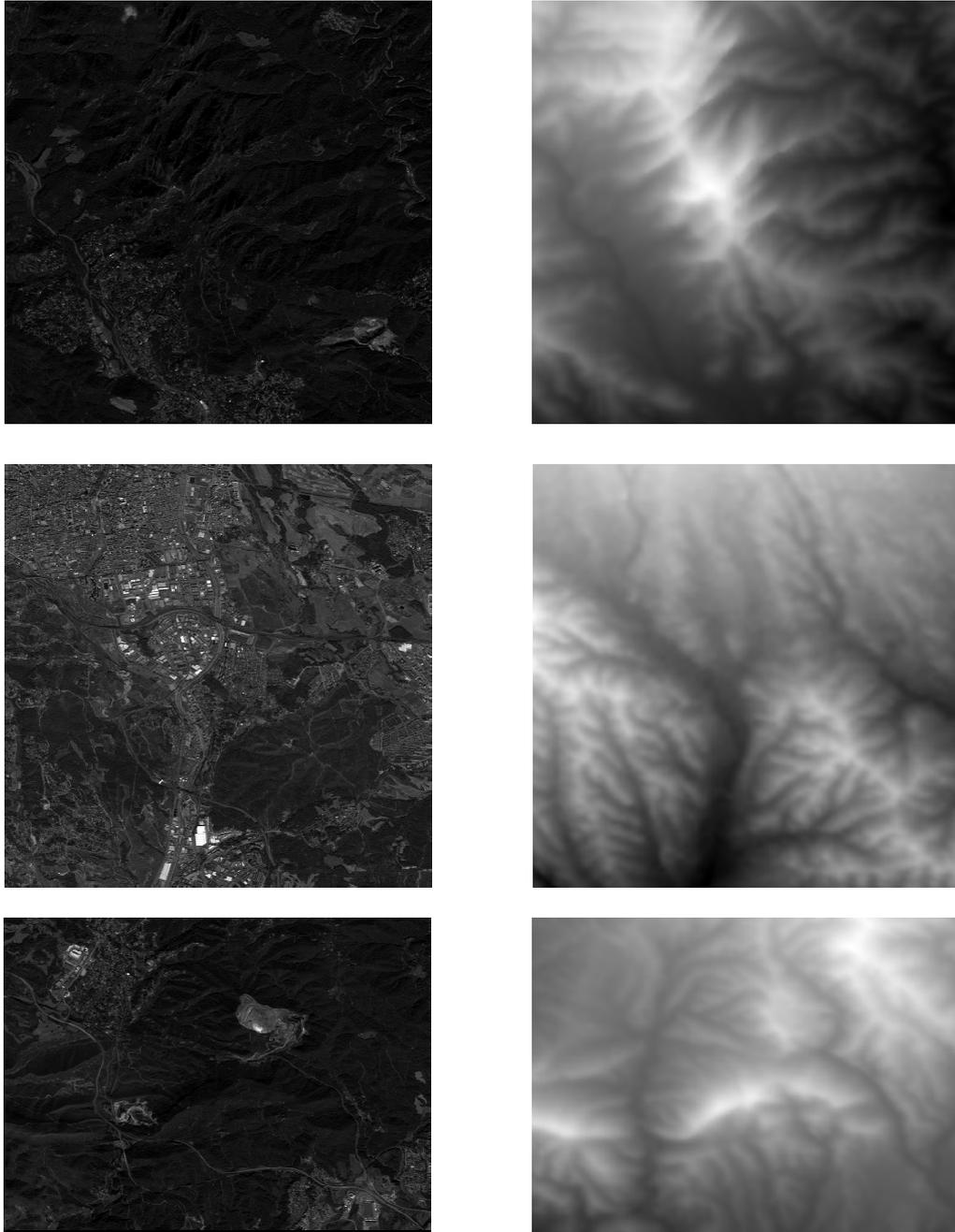


Figure 7.13. SRTM registration results for WorldView-1 Images: La Mola (first row), Terrassa (second row) and Vacarisses (third row) reference images (enhanced for better display)

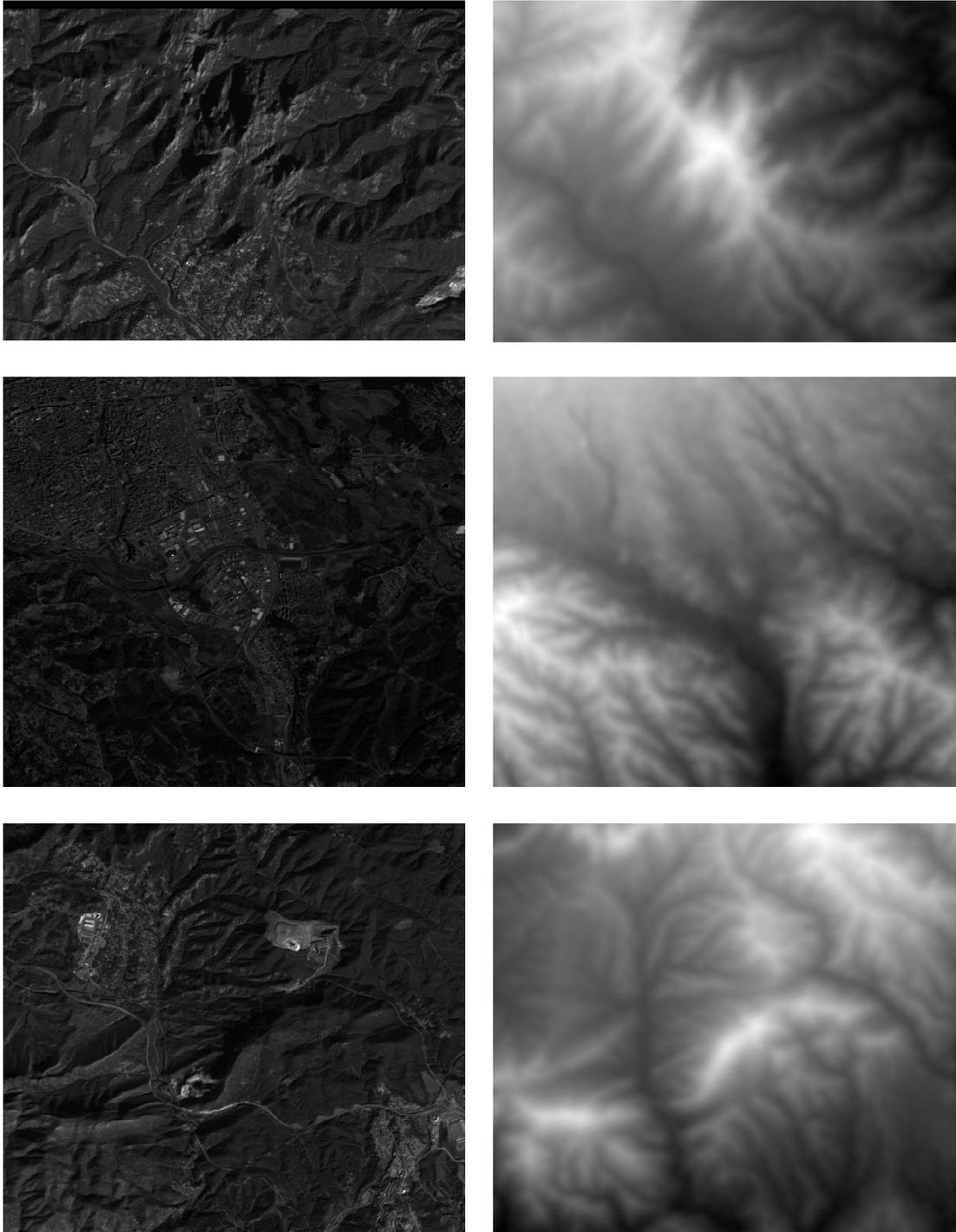


Figure 7.14. SRTM registration results for Cartosat-1 Images: La Mola (first row), Terrassa (second row) and Vacarisses (third row) reference images

7.5 Sparse Reconstruction

7.5.1 Correspondence Generation

Numerical performance evaluation for correspondence generation is a difficult task for spaceborne stereo images due to the unavailability of the ground truth and excessive number of correspondences. For example, for the IKONOS stereo pair, more than 7 million pairs were generated. The best method for performance assessment is visual inspection for various types of regions.

The accuracy of the pairs has a direct effect on the generated DSM. Therefore, the success in DSM generation can also be linked to the feature extraction process.

In this section, various patches are selected to demonstrate the matching performance. The selected patches are known to be characteristic in spaceborne images: Forests, urban, sub-urban and industrial zones. Typically, the urban zones are problematic for the photogrammetry studies, due to the occlusions caused by tall buildings and the reflectance differences in the stereo pair, especially for the roofs. The forest regions also cause significant problems due to their repeated patterns and significant visual differences between the stereo image pair: very close trees with different heights result in different image patches in the two images due to the large B/H ratio.

During the experiments, it was observed that, the RPC bias equalization scheme provided a significant improvement on the initial estimate locations, reducing the search zone down to a few pixels for points on the ground and the points in dense vegetation areas (such as forests).

KLT on Harris Features

KLT on Harris features was used to generate very accurate sub-pixel matches, but the number of correspondences was not satisfactory. However, the results of this approach are useful in SRTM bias reduction and RPC bias equalization which require a moderate number of feature points. Therefore, the experiments on Harris-KLT combination should be examined considering the accuracy and the speed, not the number of generated correspondences.

In Figure 7.15, samples from the best 100 KLT pairs are displayed. The majority of the match regions were similar to the top-left patch of the figure. Almost all of these patches lied on the ground in smooth regions where the elevation variance was very small.

The Harris-KLT matching generated about 93,977 correspondences in the IKONOS Hobart image; therefore, numerical accuracy assessment with visual control was difficult. Instead, the assessment of the matches can be done on the correspondences with the best and the worst 200 match scores. Such an inspection showed that, the best-score 200 correspondences were all correct and among the worst-score 200 correspondences, 17% were false. Therefore, increasing the number of matches by loosening the matching criteria continues to generate many accurate correspondences. Still, Harris-KLT combination could not generate sufficient number of correspondences. The number of feature points saturated at 1.3 million for the Hobart image.

The 2-way (forward-backward) consistency check reduced the number of correspondences by 33%, eliminating many correct correspondences. Therefore, the consistency check was used only for generating stereo correspondences that were used to compute the correction terms of SRTM registration bias removal and RFM bias equalization.

The KLT worked well in open areas, but the performance in the occlusion zones was poor. Few correspondences could be generated in these areas.



Figure 7.15. Sample patches for the best 100 KLT pairs



Figure 7.16. Results of the Harris-KLT approach in an industrial area, IKONOS, Hobart. The total number of correspondences is 1.3M

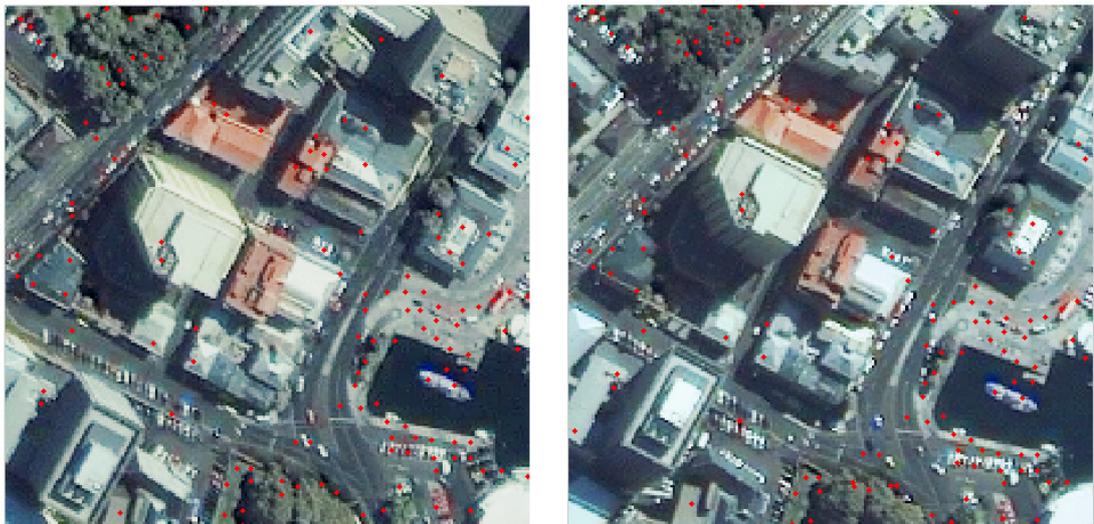


Figure 7.17. Results of the Harris-KLT approach in an urban zone, IKONOS, Hobart. The total number of correspondences is 1.3M

As it can be observed in these examples, the Harris-KLT approach provided accurate but few correspondences. Unfortunately, for spaceborne DSM generation, the accuracy is not sufficient for a correspondence

generation algorithm to be useful. The number and density of the correspondences are also important for the resultant DSMs, since using a small number of feature points causes loss of detail (Figure 7.17).

KLT on GFTT

KLT on GFTT was observed to generate many accurate correspondences in ALOS and IKONOS images, except for the occlusion zones. The results are presented in the figures below. Although both GFTT and KLT work on gray level images, the correspondences are presented here on colored images, for better assessment.

As seen in the figures, the Norikura data is challenging for any matching algorithm. KLT on GFTT was the only algorithm that succeeded in matching, but there were many “no-match” regions, especially in snow-covered areas. Snow cover is a typical unresolved problem in feature matching for spaceborne DSM generation.

In general, GFTT-KLT algorithm generated the largest number of correspondences, most of which were accurate. Thus, this algorithm is a good choice for the correspondence generation phase of a DSM generation workflow.

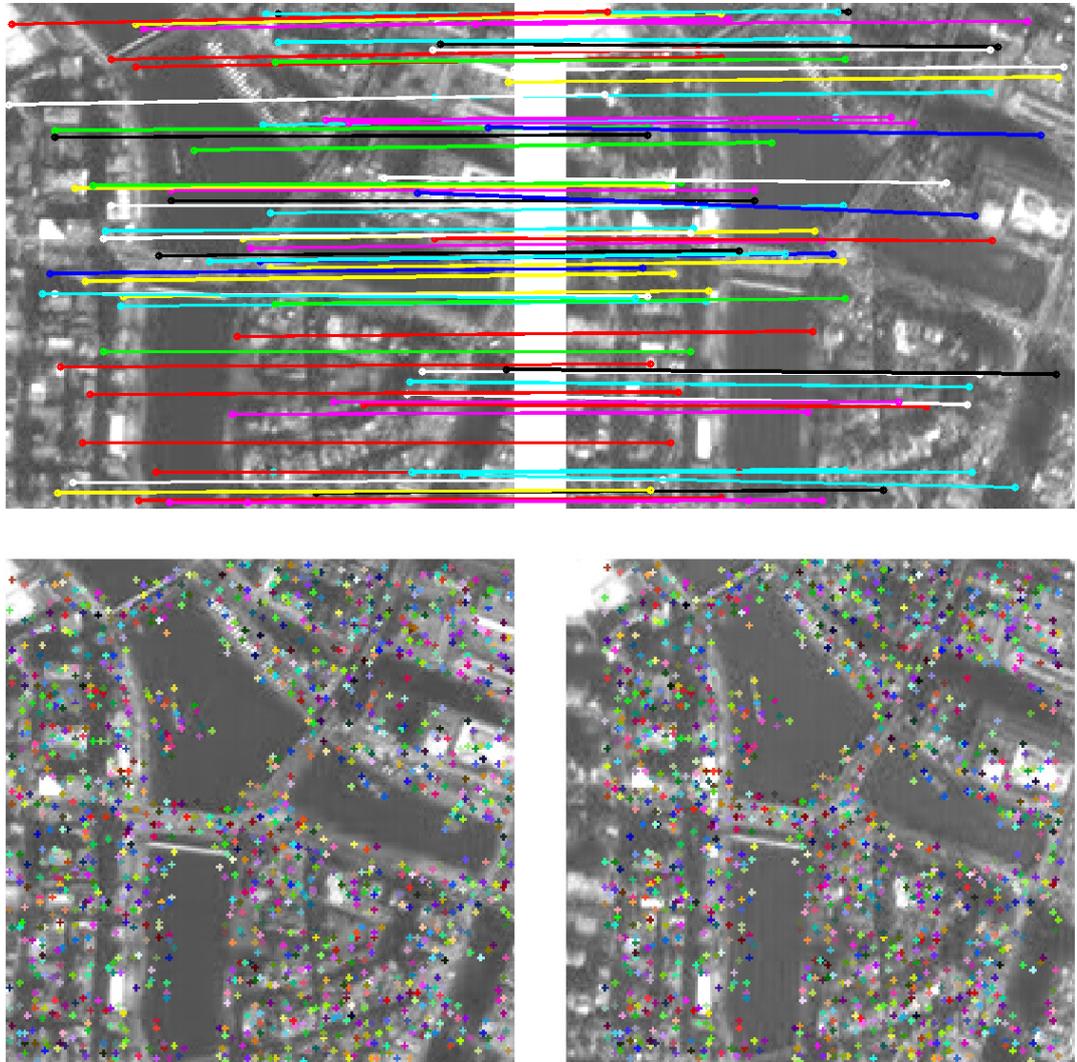


Figure 7.18. Matching results for GFTT-KLT in ALOS Hiroshima images: The needlegram (first row) displays 1/20 of the correspondences. The second row displays all of the correspondences.

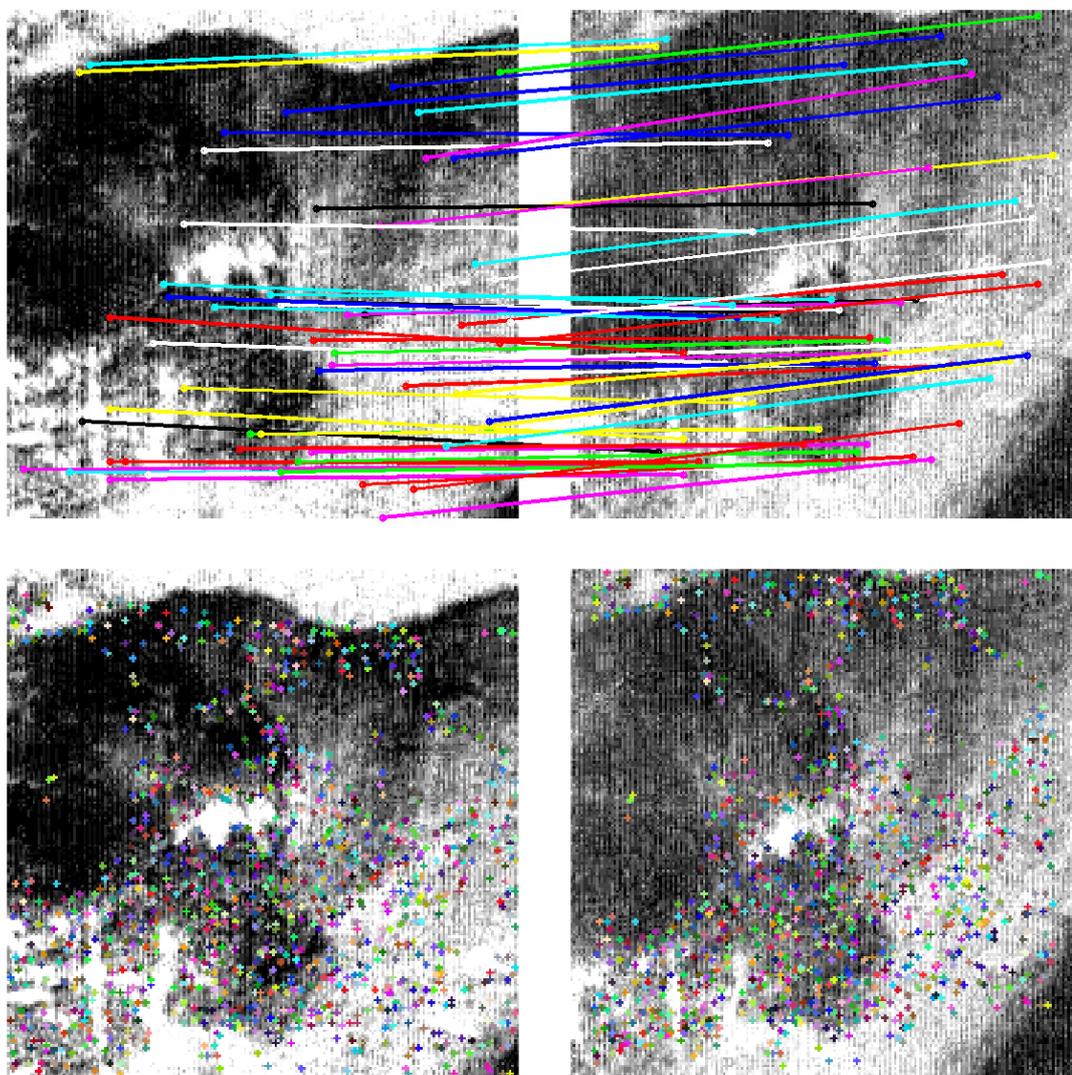


Figure 7.19. Matching results for GFTT-KLT in ALOS Norikura images: The needlegram (first row) displays 1/20 of the correspondences. The second row displays all of the correspondences.

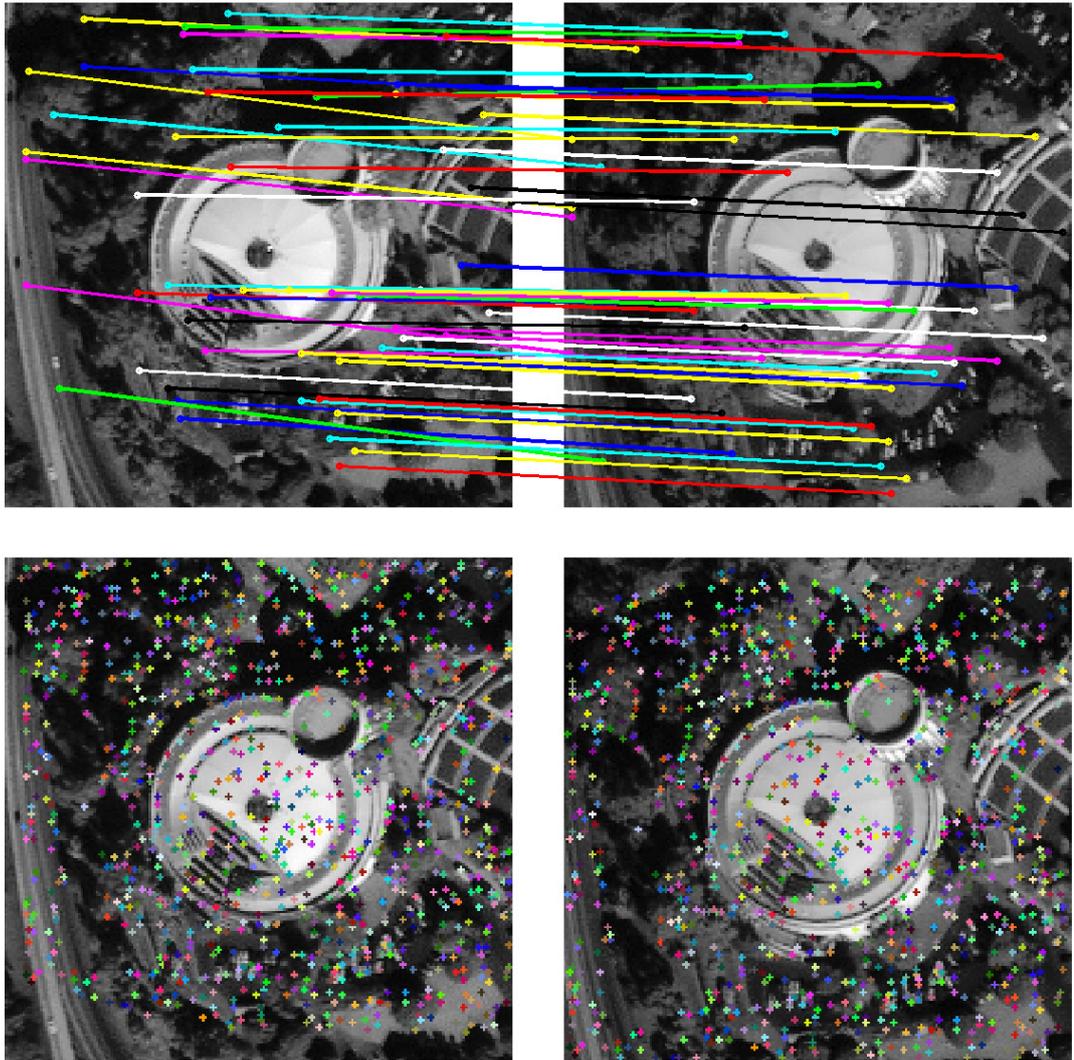


Figure 7.20. Matching results for GFTT-KLT in IKONOS San Diego images: The needlegram (first row) displays 1/20 of the correspondences. The second row displays all of the correspondences.

BRIEF + Template Matching on FAST Points

FAST-BRIEF+TM is a hybrid method which benefits from the speed of feature-based matching and the accuracy of area-based matching. Sample results of the algorithm are presented in the figures below.

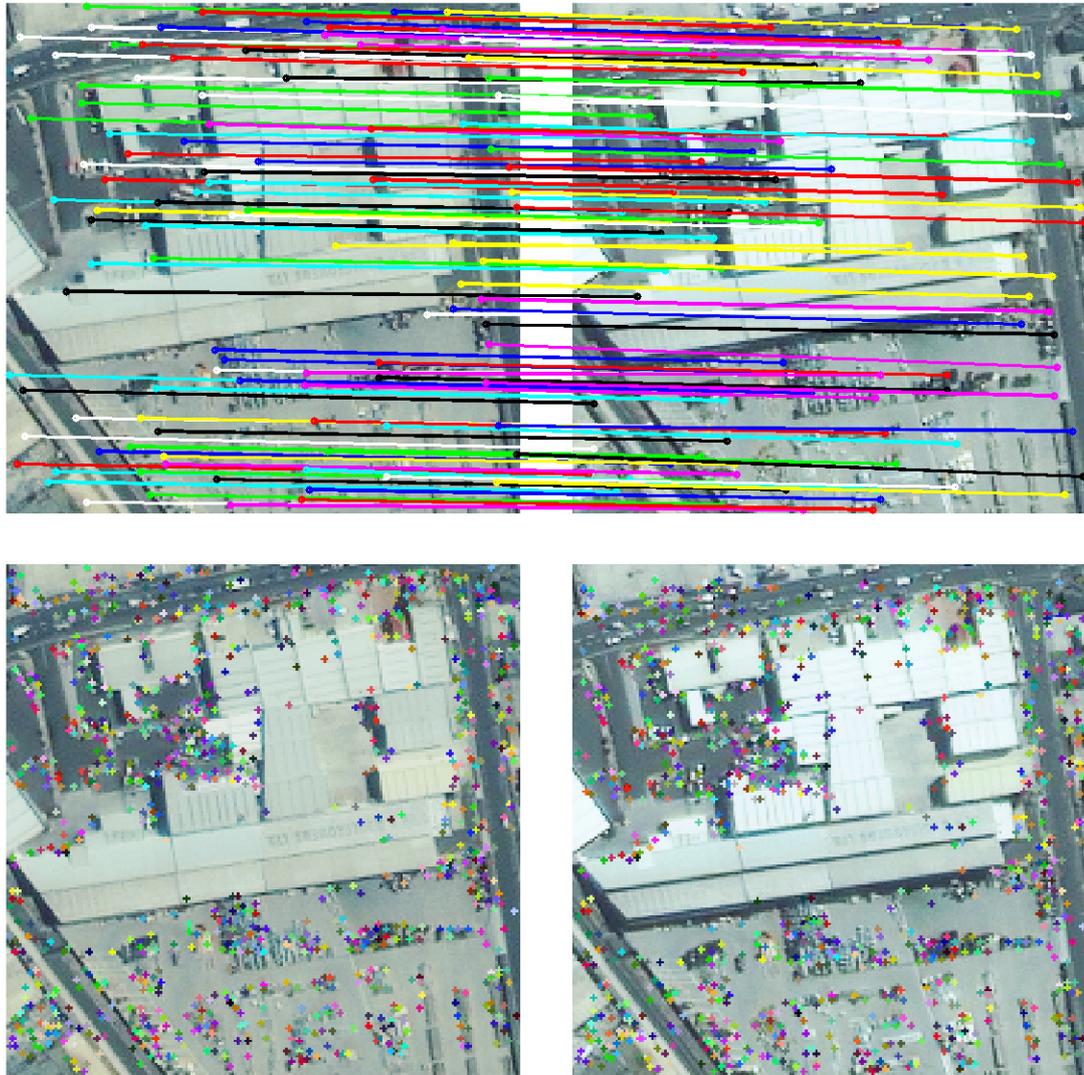


Figure 7.21. The matching results for BRIEF+TM on FAST points in an industrial region (IKONOS Hobart). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

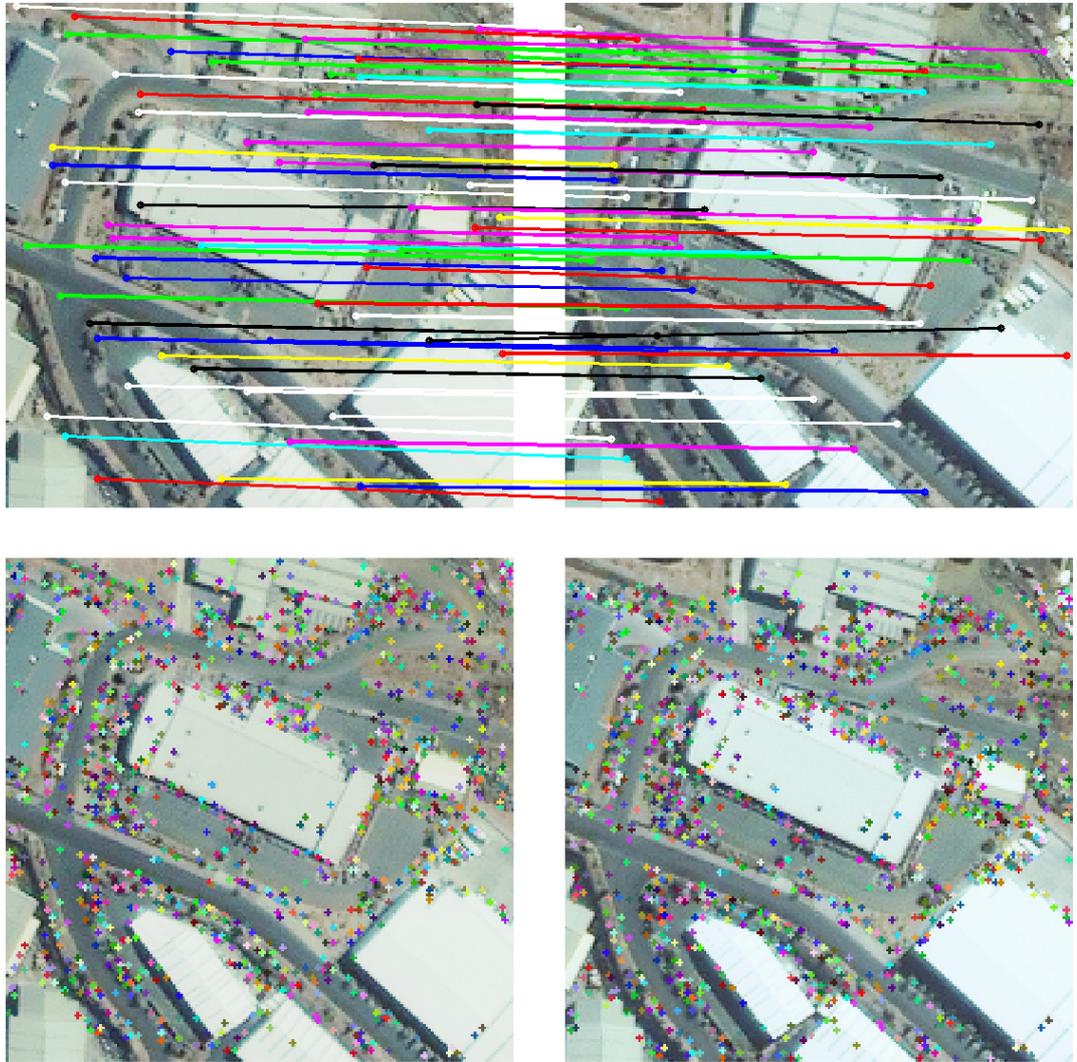


Figure 7.22. The matching results for BRIEF+TM on FAST points in an industrial region (IKONOS Hobart). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

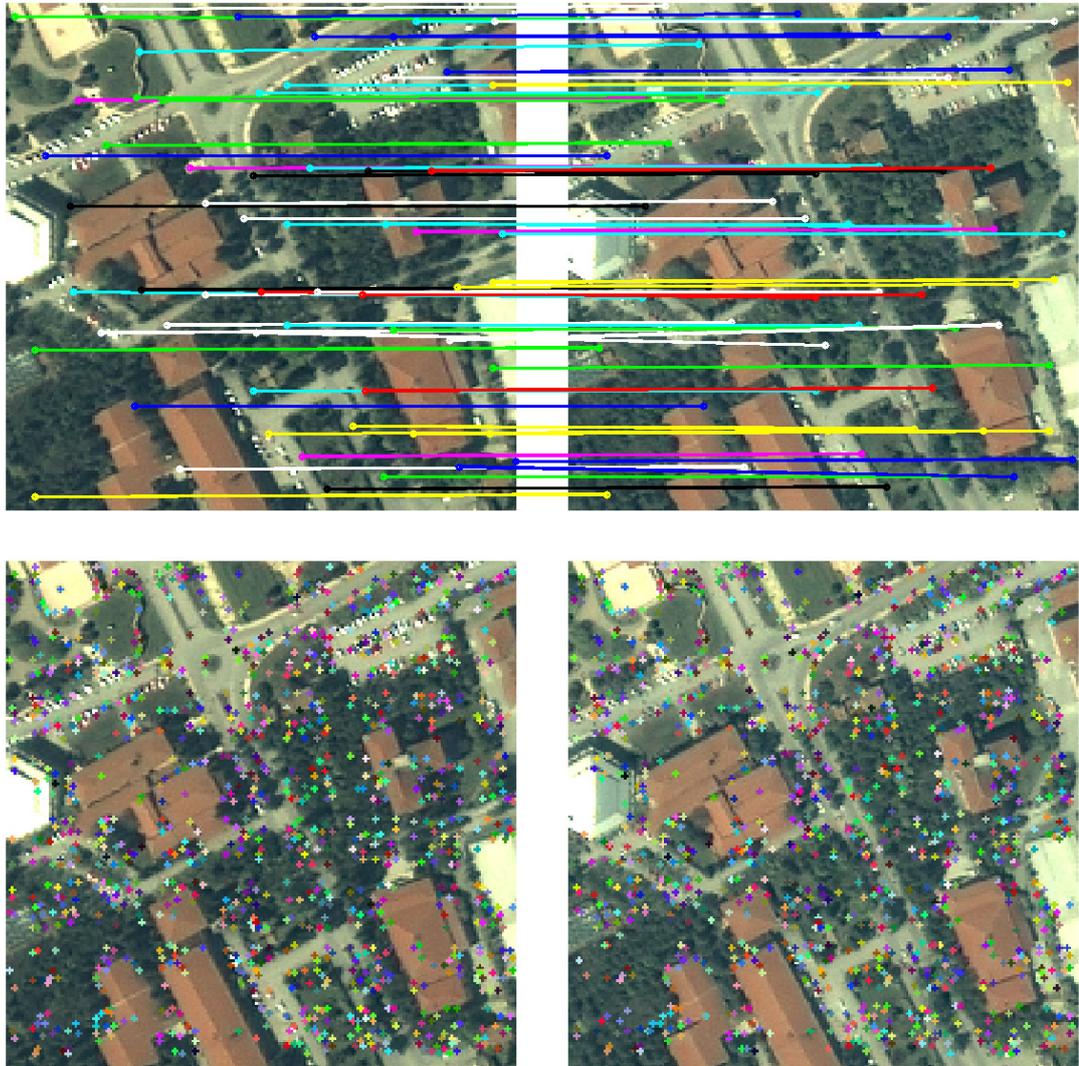


Figure 7.23. The matching results for BRIEF+TM on FAST (IKONOS Eskişehir). The first row is the needlegram for 1/10 of the correspondences. The second row displays all of the correspondences.

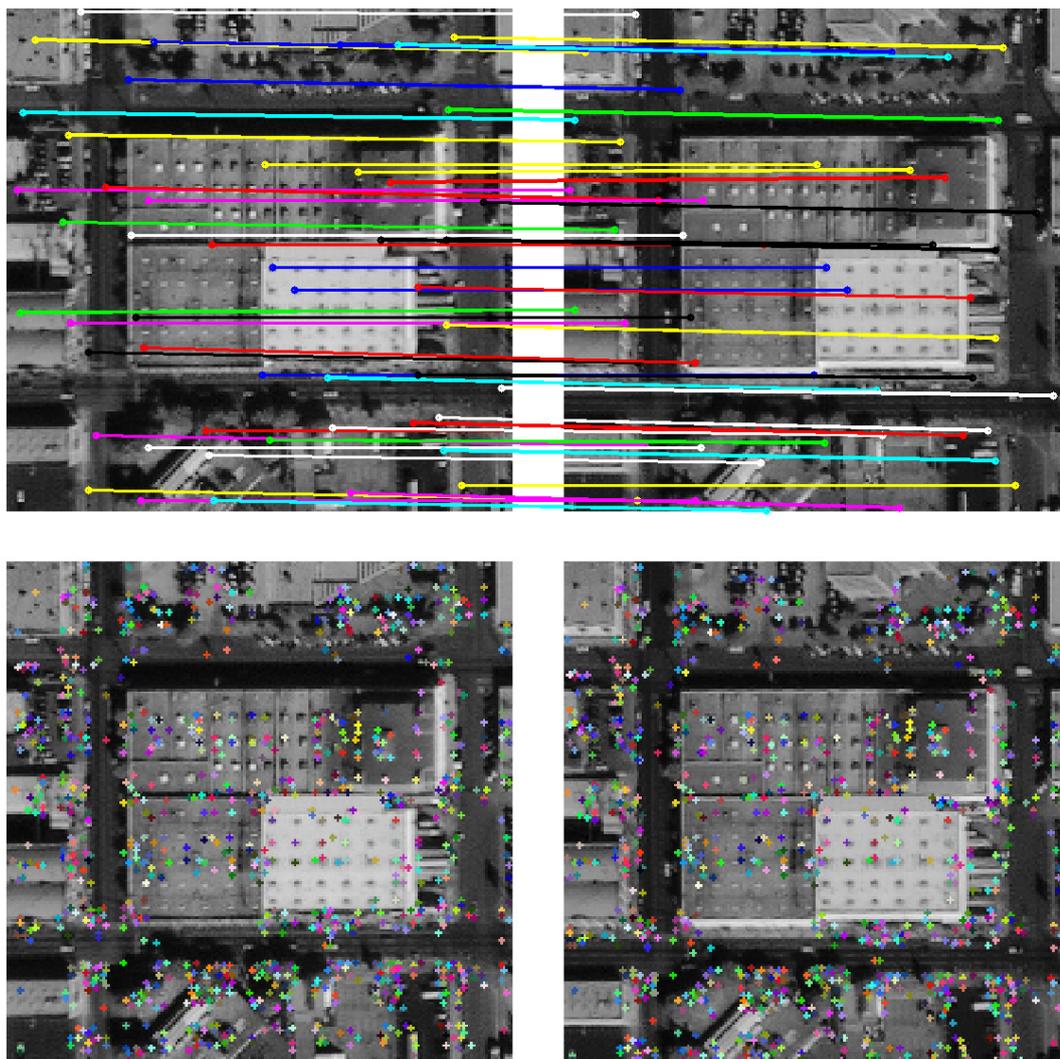


Figure 7.24. The matching results for BRIEF+TM on FAST points in an urban region (IKONOS San Diego). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

It is difficult to assess the matching performance in forests, even for the human eye. Here, we present a relatively large patch to present a clue for the matching performance (Figure 7.25): The FAST points are detected on grids to ensure uniform distribution in the image. The grid based detector uses non-overlapping patches, and causes the no-feature strips (5-pixel thick) in image 1. For image 2, FAST points are also detected by a grid-based detector,

again on a regular grid. However, these points are moved to the correct match positions by template matching, which may displace the FAST points up to four pixels. As it can be seen in Figure 7.25, there is a distorted grid in the second image. These strips are not caused by the grid-based FAST detector running on image 2, since it is not a regular grid. Additionally, the grid boundaries in image1 and image 2 do not pass over the same ground positions. Therefore, the preservation of the empty strips (of course with deformations caused by stereo imaging) is an indication of a good matching strategy.

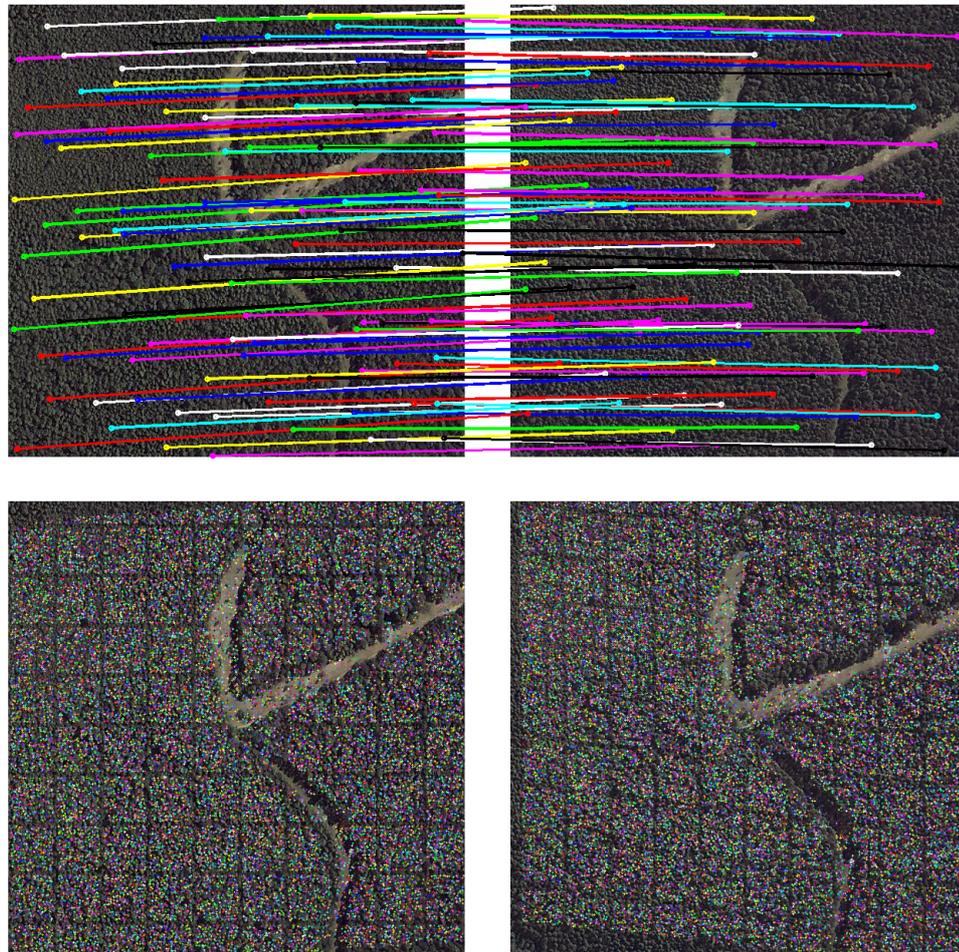


Figure 7.25. Matching results for BRIEF+TM on FAST points in forest (IKONOS Hobart). The first row is the needlegram for 1/200 of the correspondences. The second row displays all of the correspondences.

Typically, forests are known to be problematic in correspondence generation for spaceborne imagery. However, the proposed algorithm is very successful in these regions. The reason for such success lies in the usage of SRTM registration. Classical photogrammetric approaches utilize coarse DSMs obtained using hierarchical methods. However, the case in higher levels of the image pyramid is even worse for matching in the forest regions, thus the algorithms fail. When SRTM registration is used to constrain the search region, tall image pyramids are not required and the probability of encountering a false match with a similar appearance is reduced. Infrequent false matches may occur, but they are eliminated with the neighbor consistency check.

Observed Problems

The sparse matching step generates many false matches. The used matching strategy is designed specifically to generate an excessive amount of correspondences. The underlying idea in this strategy is the assumption that; statistically, false matches are scattered among the correct matches (which constitute the great majority) and thus in the interpolation step, a false match is expected to influence only a small area. This strategy provides relatively good results in the forest areas, where the density of correspondences is very high.



Figure 7.26. Grouped false matches (FAST-BRIEF+TM) due to occlusions in urban areas (IKONOS Hobart). A group of false matches is marked with red ellipses. Matching errors due to difference in the reflectance are marked with a white ellipse

Grouped false matches are rare but existent, especially in the occluded areas of the urban regions (Figure 7.26). Such matches cannot be eliminated by neighbor features' consistency (described in Section 5.2.3), since the neighbors are also false matches. This problem is unavoidable during the matching phase. Further elimination is performed during the reconstruction phase by thresholding the projection error for the reconstructed object point. Moreover, some of the interpolation algorithms reduce the effect of outliers.

Eskişehir images cover mostly an urban area, where there exist adjacent and isolated tall buildings, which cause significant occlusions. Matching features on tall building rooftops require wide search regions and hierarchical algorithms are utilized for that purpose. However, the occlusion zones misguide these algorithms in higher levels of the image pyramid, causing false correspondences, most of which are eliminated by the

consistency check. The resulting correspondence image has large “no-match” regions or false matches in these regions. The problem cannot be resolved by using non-hierarchical search methods due to the large search zone: When small templates are used, the matching algorithm often finds a false match with a similar (but false) window; when the templates are large, the outcome resembles that of the hierarchical methods.

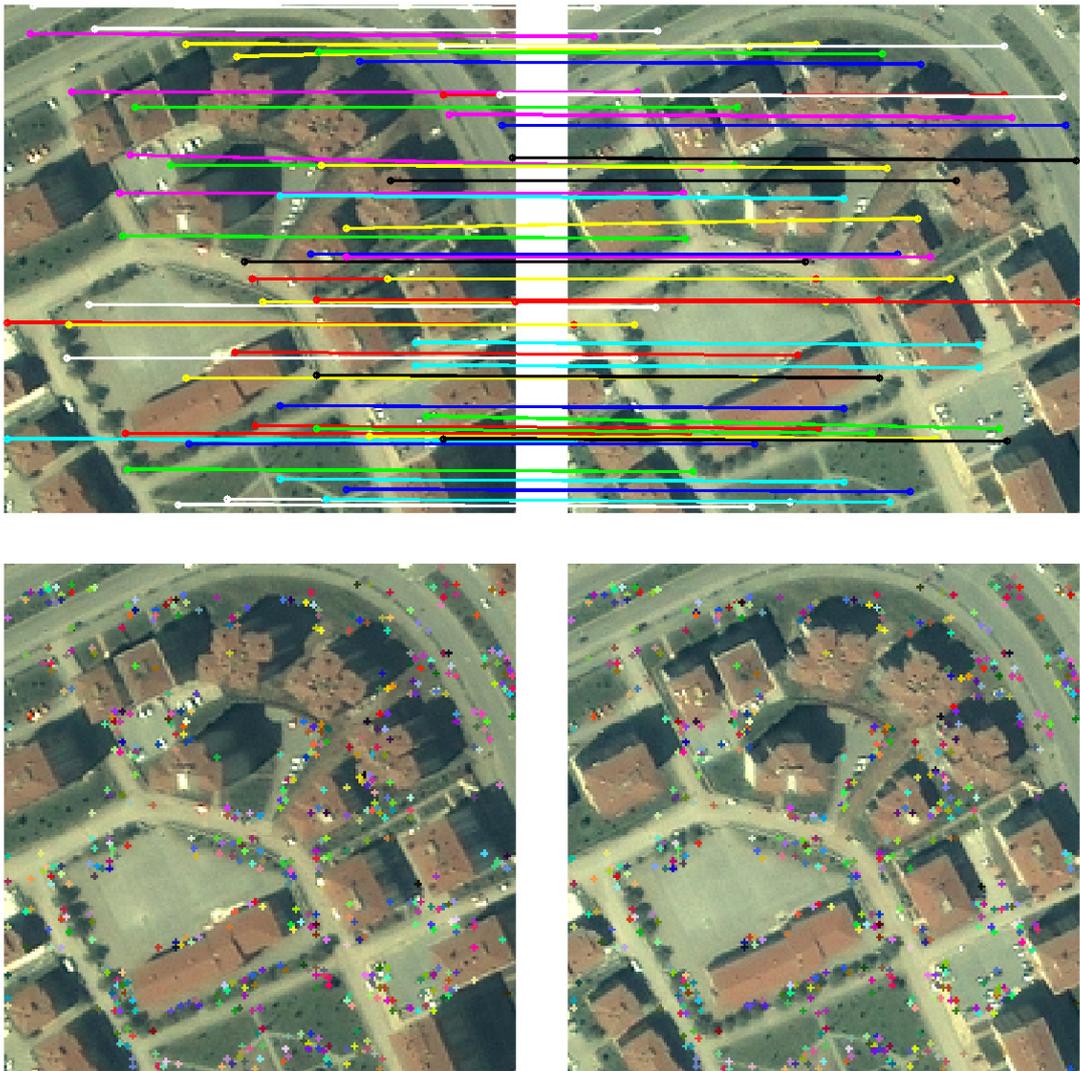


Figure 7.27. Problems caused by occlusions due to isolated tall buildings for BRIEF+TM on FAST points (IKONOS Eskişehir). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

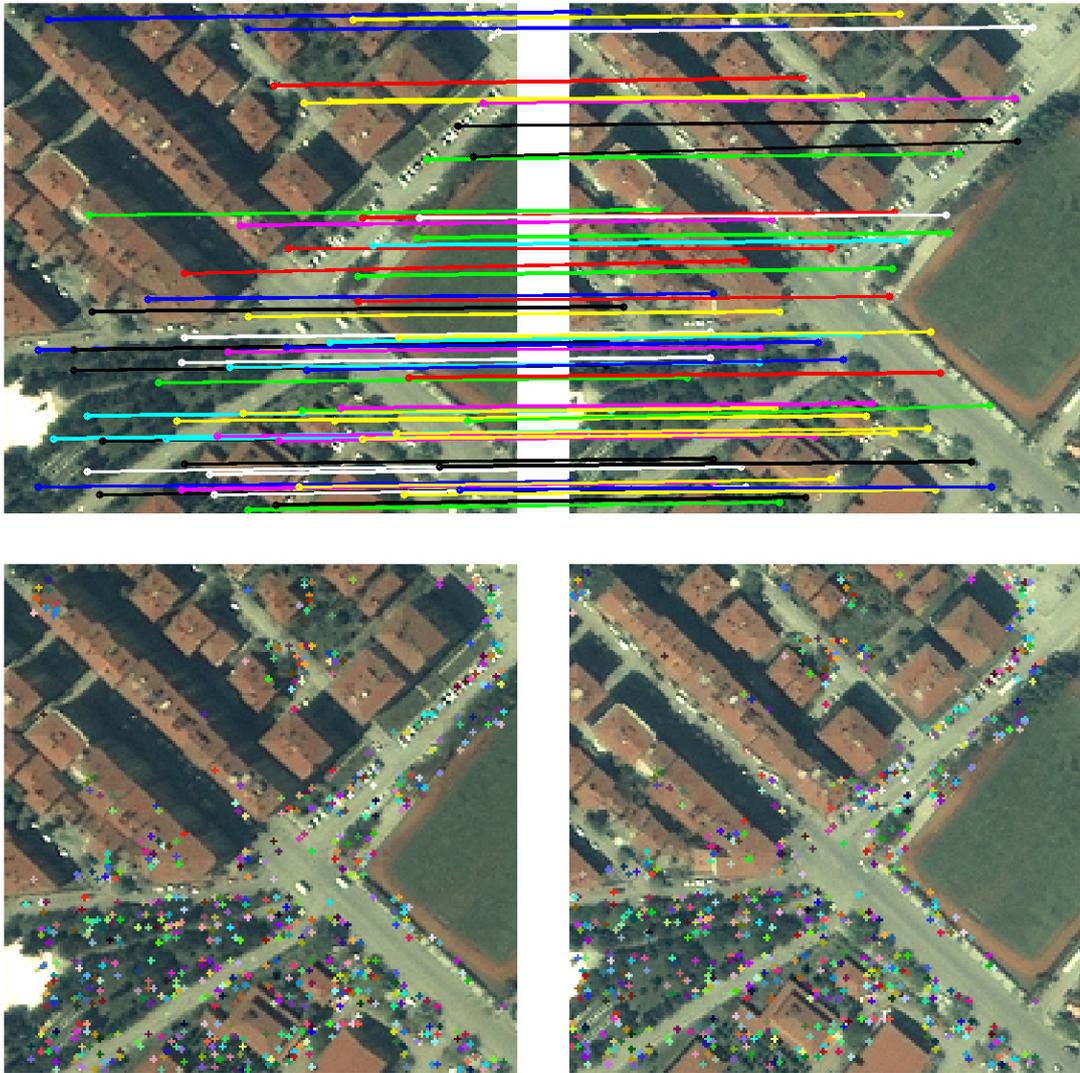


Figure 7.28. Problems caused by occlusions due to adjacent tall buildings for BRIEF+TM on FAST points (IKONOS Eskişehir). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

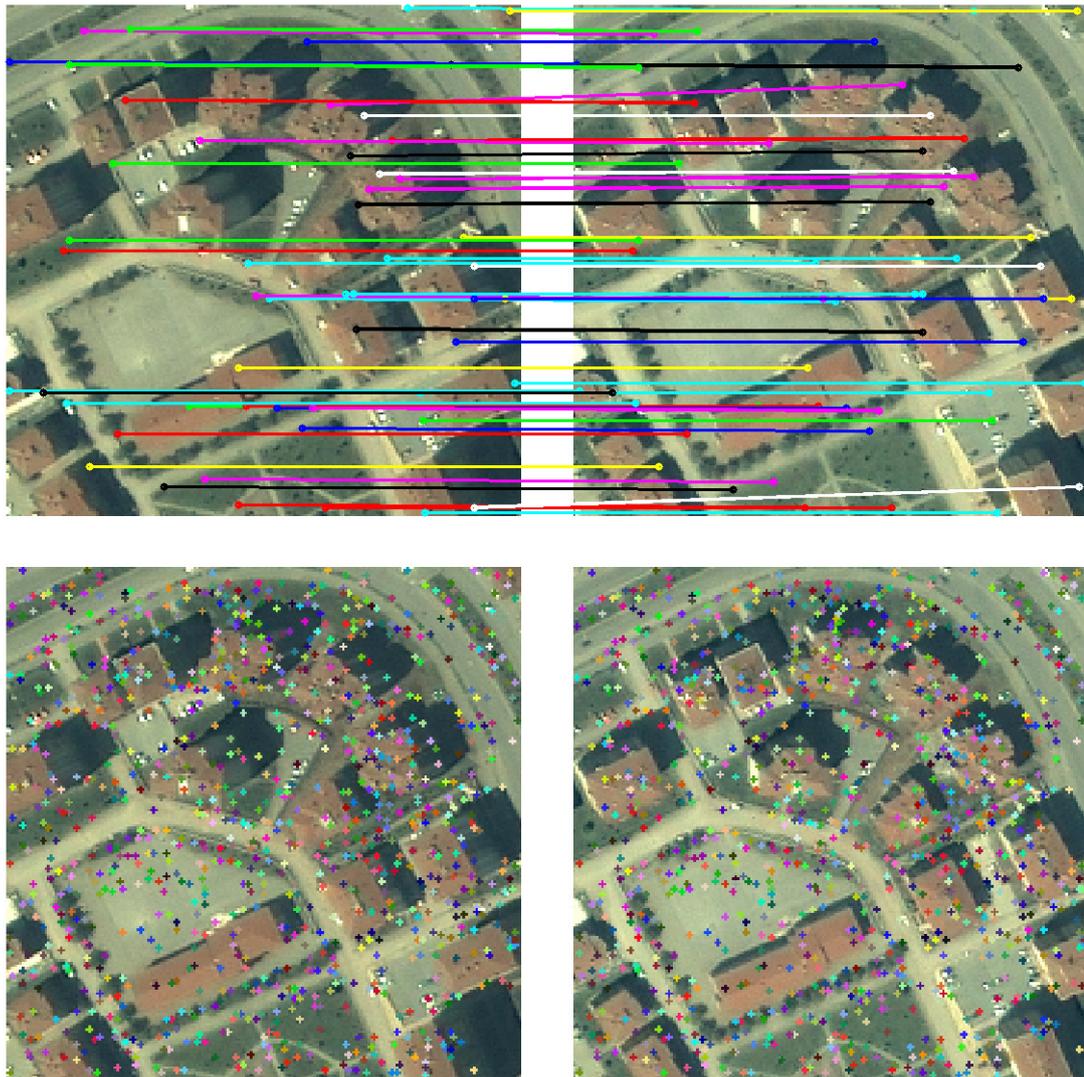


Figure 7.29. Problems caused by occlusions due to isolated tall buildings for KLT on GFTT points (IKONOS Eskişehir). The first row is the needlegram for 1/20 of the correspondences. The second row displays all of the correspondences.

GFTT-KLT algorithm generates a large number of correspondences almost everywhere. This behavior may cause errors in the occlusions zones, where the majority of the correspondences are false. In these regions, the neighbor elimination algorithm fails to perform a successful elimination and false positioned correspondences are passed to the dense DSM generation

algorithms causing invalid results. However, this problem can be overcome by reducing the number of correspondences with a higher quality threshold for the features.

The proposed matching algorithms fail to generate usable correspondences in WorldView-1 and CARTOSAT-1 images. The main reasons are: *i)* scale changes between the images, *ii)* significant occlusions and appearance changes *iii)* geolocation error in the SRTM data, *iv)* steep terrain and *iv)* radiometric inconsistency between the stereo images (see Section 7.1). The proposed matching strategy can overcome radiometric inconsistencies or appearance changes to a certain extent, and scaling can be corrected manually. However for the WorldView-1 and CARTOSAT-1 images, the changes in appearance and geometric deformations are radical. The half-pixel geolocation error in the SRTM data causes 80 pixels of registration error for the WorldView-1 images and the succeeding matching algorithms require large search ranges, in which false matches are unavoidable under the above-mentioned deformations. For the CARTOSAT images, the provided RPCs have 150 to 250m error on the ground that causes significant errors in the SRTM registration. Since the region of interest has fast changes in terrain height, the registration accuracy is important for obtaining proper initial estimates. Therefore, the registration output cannot be used to constrain the search zone for the matching algorithm.

7.5.2 Reconstruction

RPC reconstruction was implemented as explained in Chapter 5 and Appendix B. The Levenberg-Marquardt algorithm was fed with the initial estimates obtained from the SRTM registration. In all cases, less than 10 iterations were sufficient for convergence.

The reconstructed point was discarded from the point list, if its total re-projection error (sum of errors in both images) was more than 3 pixels. Such points exist and are mostly caused by errors in matching. This elimination method reduced the total number of points by 10%. It was observed that using RPC bias equalization terms reduced the number of rejected points significantly. This is an expected result, since the equalized bias terms produces compatible projection functions.

The modified projection functions (Section 5.3.1) were tested for the improvement in reconstruction accuracy in the Hobart region, on 122 GCPs. The results are summarized in Table 7.4. As seen in the table, the error in the reconstructed height was decreased by 2 meters on the average. On the other hand, the geolocation accuracy has degraded for about 0.65×10^{-5} degrees. This corresponds to 50 cm (approximately) increase in the mean geolocation error.

Table 7.4. Reconstruction errors for original RPCs and RPCs with error correction terms that are generated using bias equalization, IKONOS Hobart

	Without RPC bias equalization			With RPC bias equalization		
	Latitude Error (degrees)	Longitude Error (degrees)	Height Error (meters)	Latitude Error (degrees)	Longitude Error (degrees)	Height Error (meters)
μ	$2,44 \times 10^{-5}$	$-4,48 \times 10^{-5}$	-3,49	$2,54 \times 10^{-5}$	$-5,12 \times 10^{-5}$	-1,46
σ	$5,53 \times 10^{-6}$	$8,41 \times 10^{-6}$	0,77	$5,53 \times 10^{-6}$	$8,42 \times 10^{-6}$	0,771

The RPC reconstruction resulted in 4.5m geolocation error for the biased RFMs and 5m geolocation error in the bias-equalized RFMs. As explained in Section 5.3.1, the main objective in bias equalization is to improve the height accuracy; an increase in geolocation error is expected. The error in geolocation is caused by the common image biases for the two images (Table 7.1). It should be noted that the standard deviations of the

reconstruction errors are small (less than 1m) and unchanged for the biased and bias-free RFMs.

7.6 Dense DSM Generation

7.6.1 Performance of the Proposed Methods

Interpolation was performed with *i)* interpolation with edge-aware filter, *ii)* interpolation with meshing, *iii)* interpolation with segmentation and *iv)* interpolation using integral images, using the sparse reconstruction results for the FAST-BRIEF+Template Match and GFTT-KLT pairs.

Accuracy assessment for the dense DSMs is not possible due to the unavailability of ground truth or more accurate data. This is a typical problem in the photogrammetry studies in which the accuracy can only be measured on sparse GCPs. Thus, visual inspection of the results should be made to observe the distinguishable objects and details, but not for accuracy comparison among the methods.

The performance of the interpolation methods that use the image data as the interpolation guide is dependent on the following factors:

- i) The number and the distribution of sparse points,
- ii) The existence of color,
- iii) Image resolution,
- iv) Noise level,
- v) Image texture.

Therefore, for testing the interpolation algorithms it is not required to present the interpolation results for all satellites. Instead, here we present the results on characteristic regions that vary for the factors affecting the interpolation performance.

Interpolation results for IKONOS Hobart stereo data are presented in Figure 7.31 to Figure 7.40. The regions are selected from urban, rural, industrial and forest areas (Figure 7.30) to provide diversity.

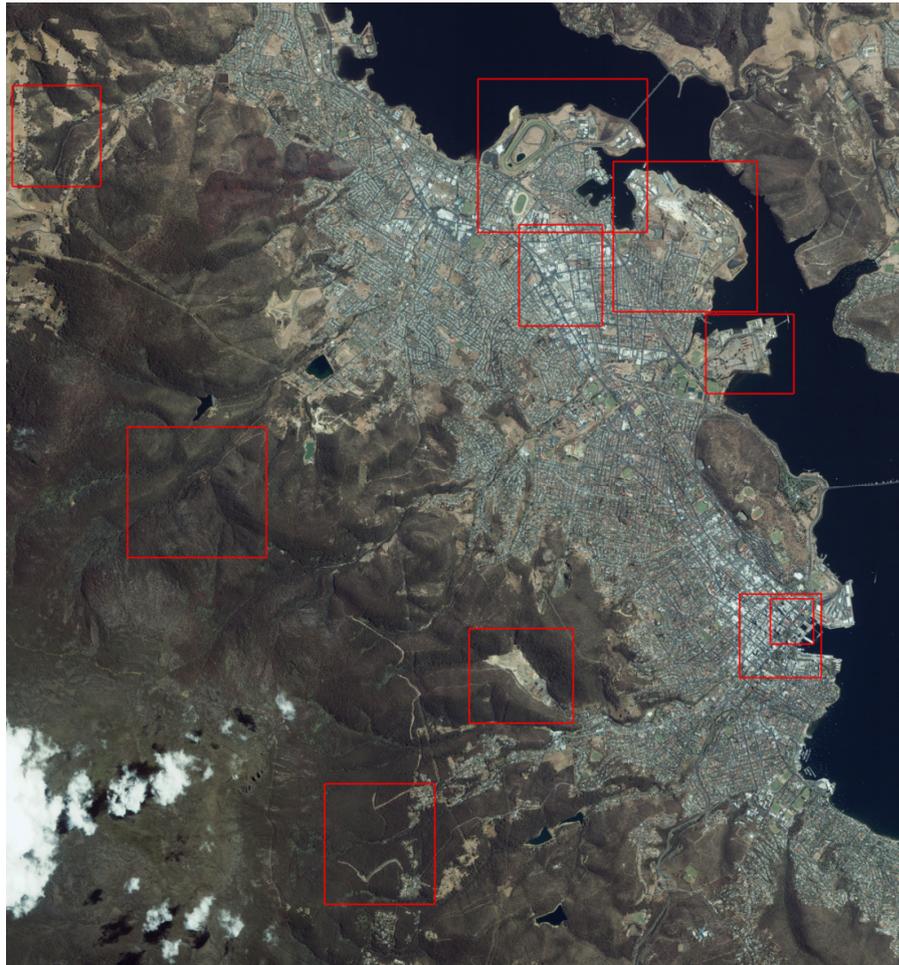


Figure 7.30. Selected regions in the IKONOS Hobart image

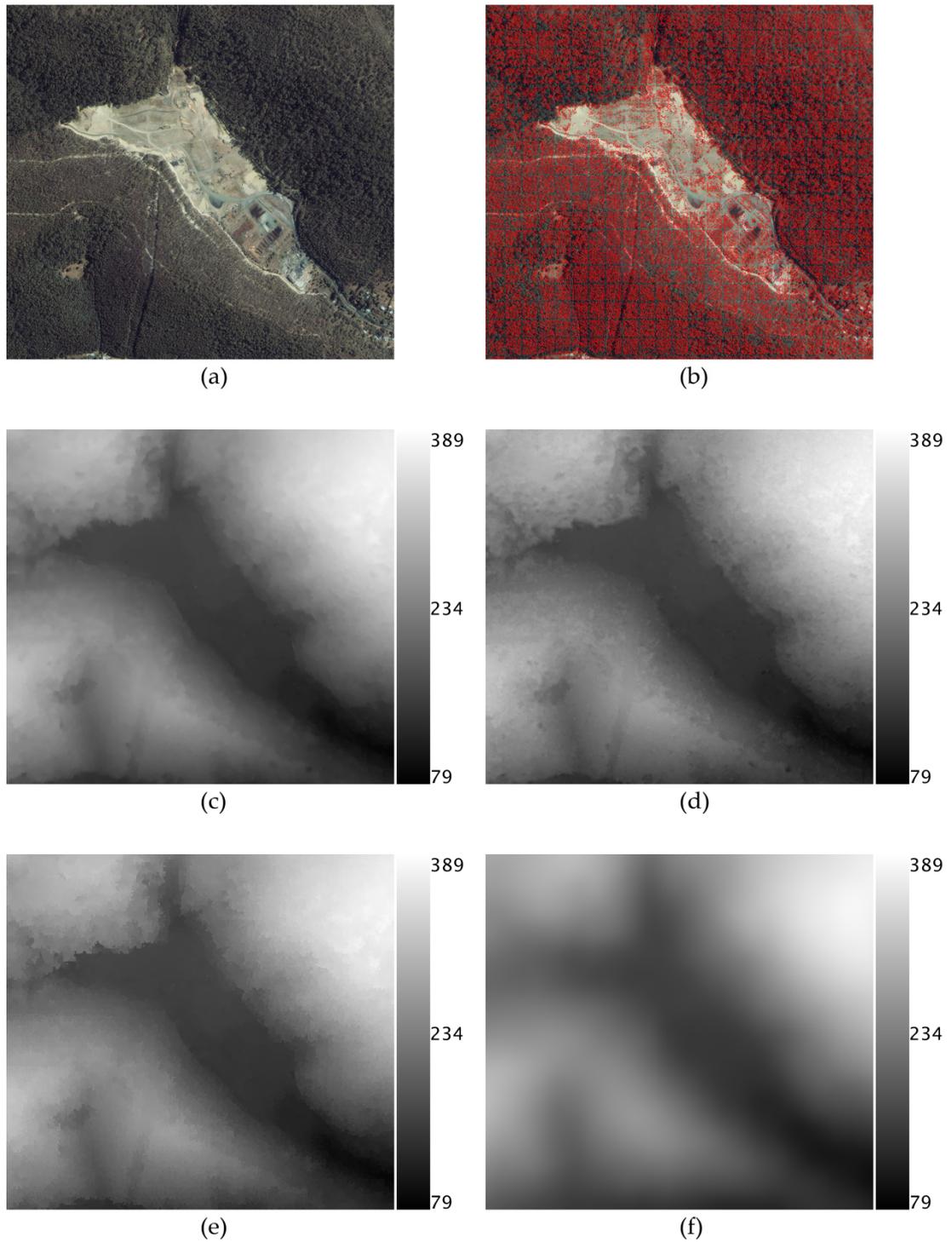


Figure 7.31. Interpolation results for various algorithms on patch 1 (1401x1290): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

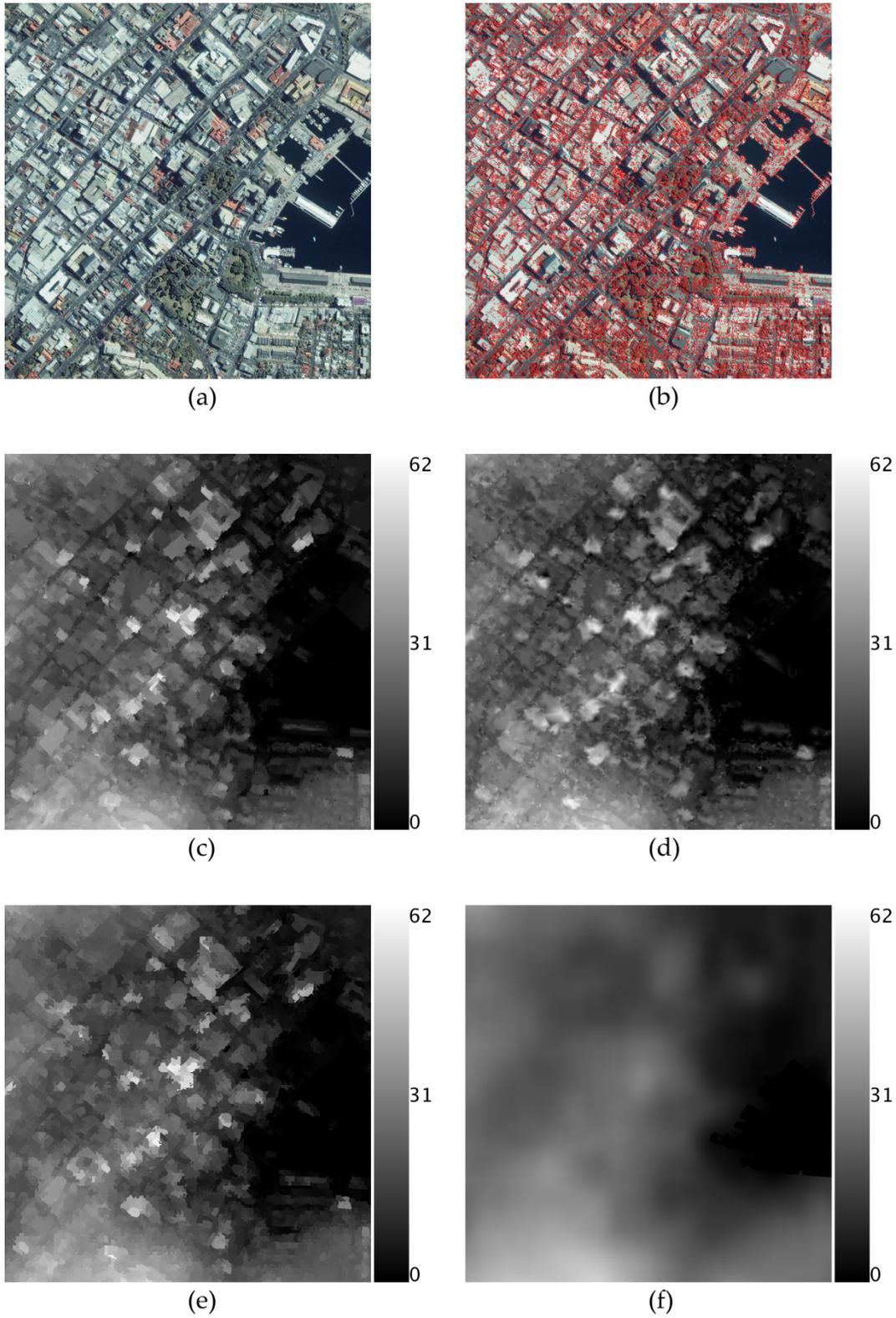


Figure 7.32. Interpolation results for various algorithms on patch 2 (1099x1138) (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

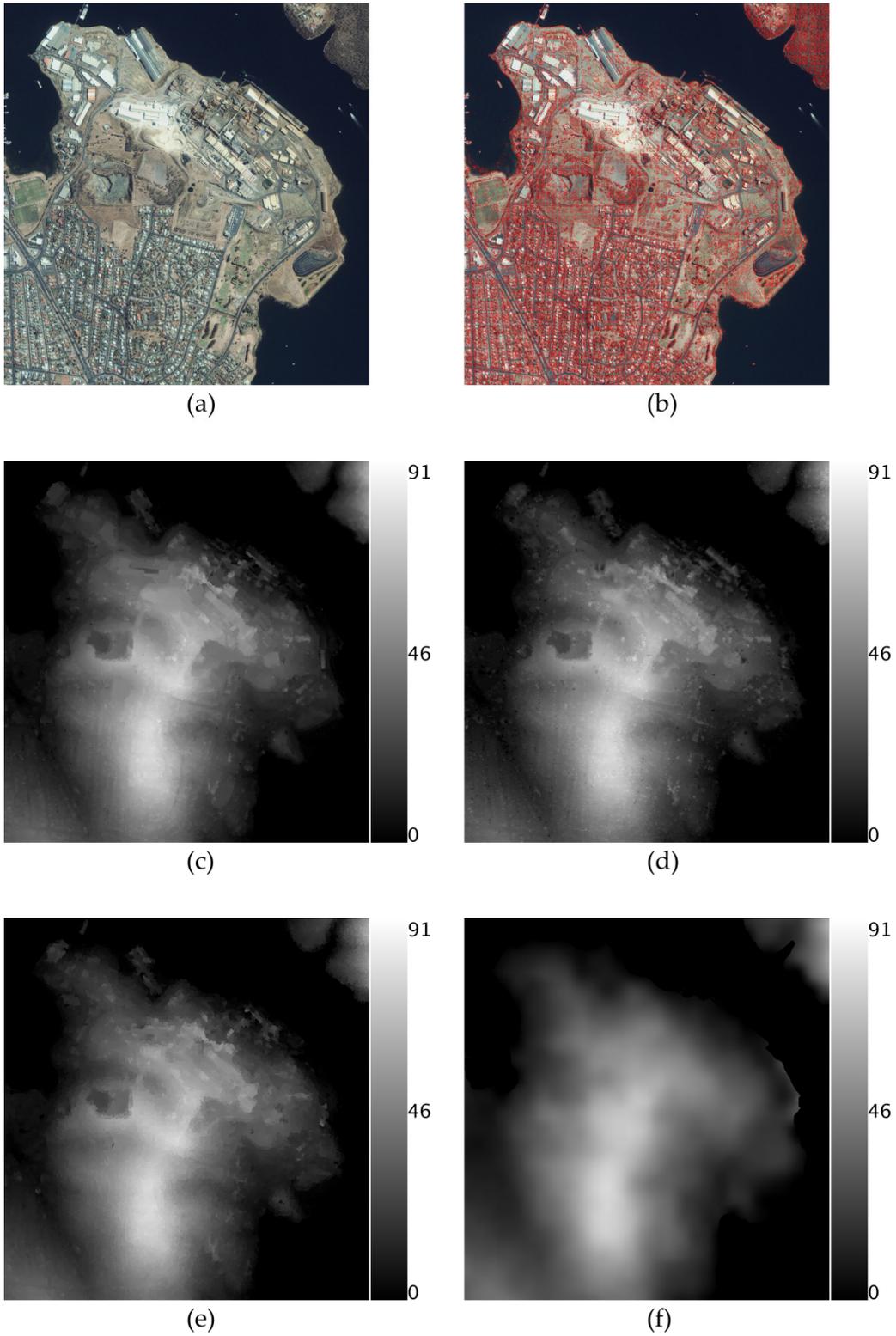


Figure 7.33. Interpolation results for various algorithms on patch 3 (1936x2041): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

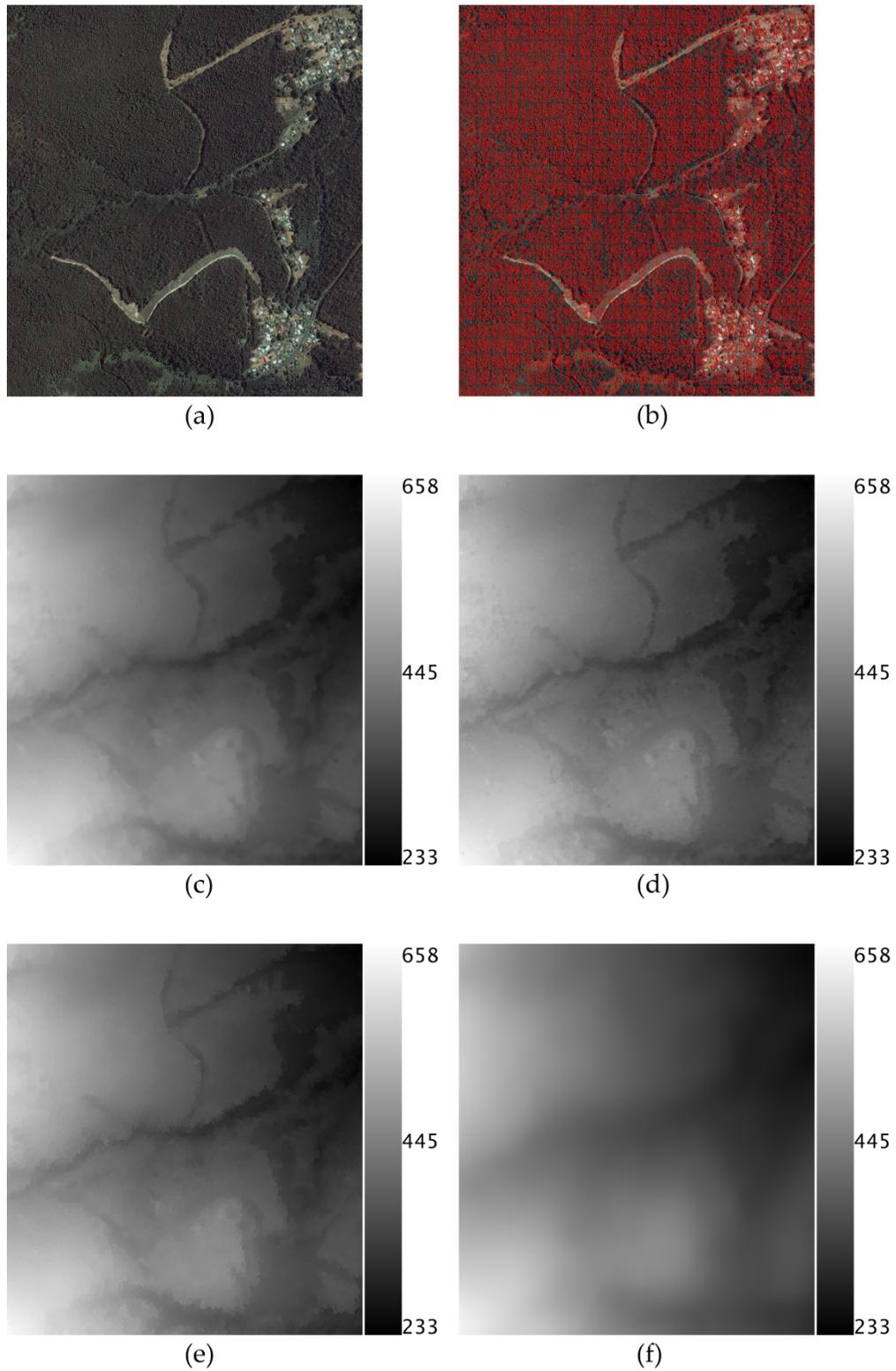


Figure 7.34. Interpolation results for various algorithms on patch 4 (1484x1642): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

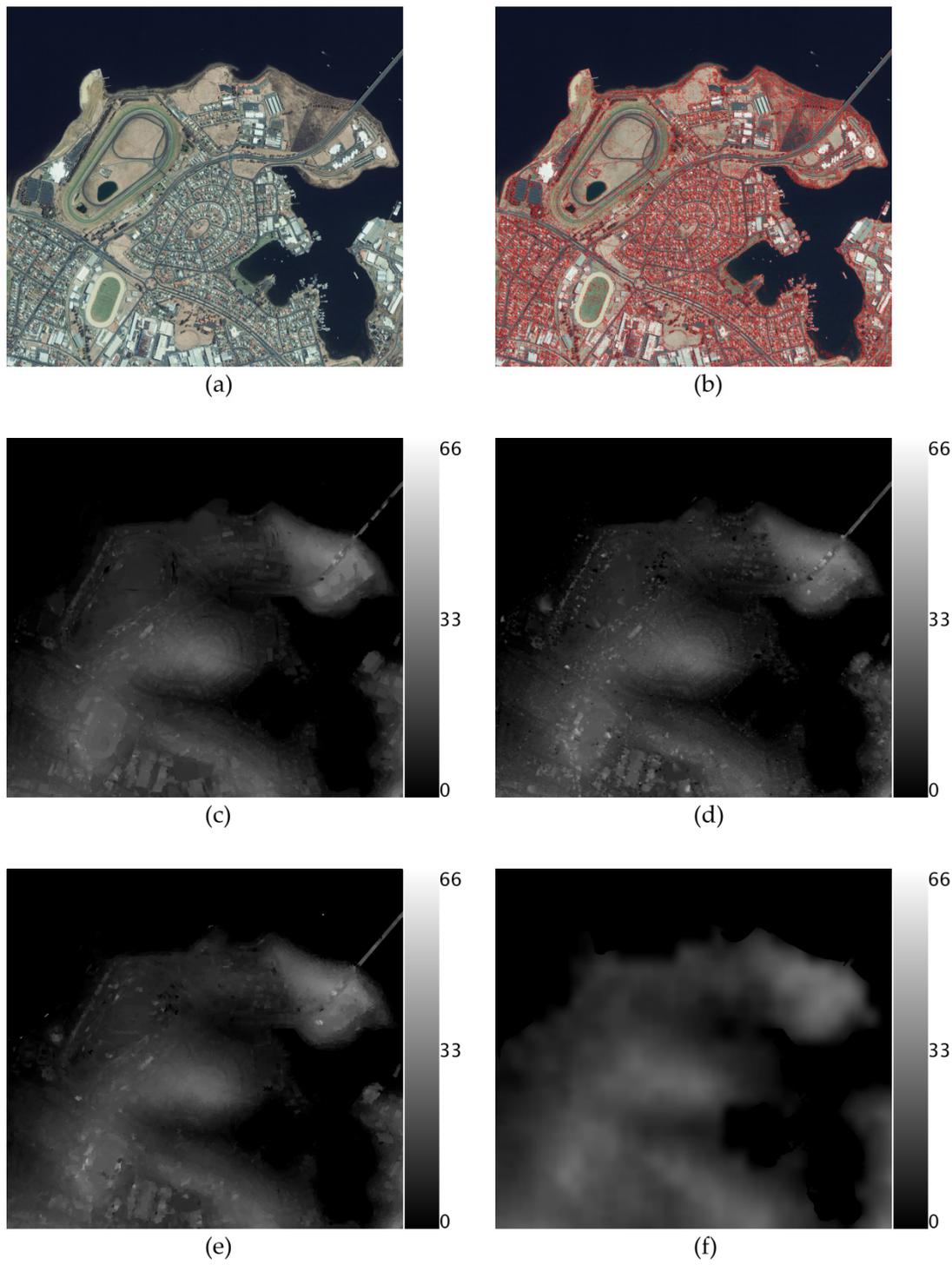


Figure 7.35. Interpolation results for various algorithms on patch 5 (2281x2083): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

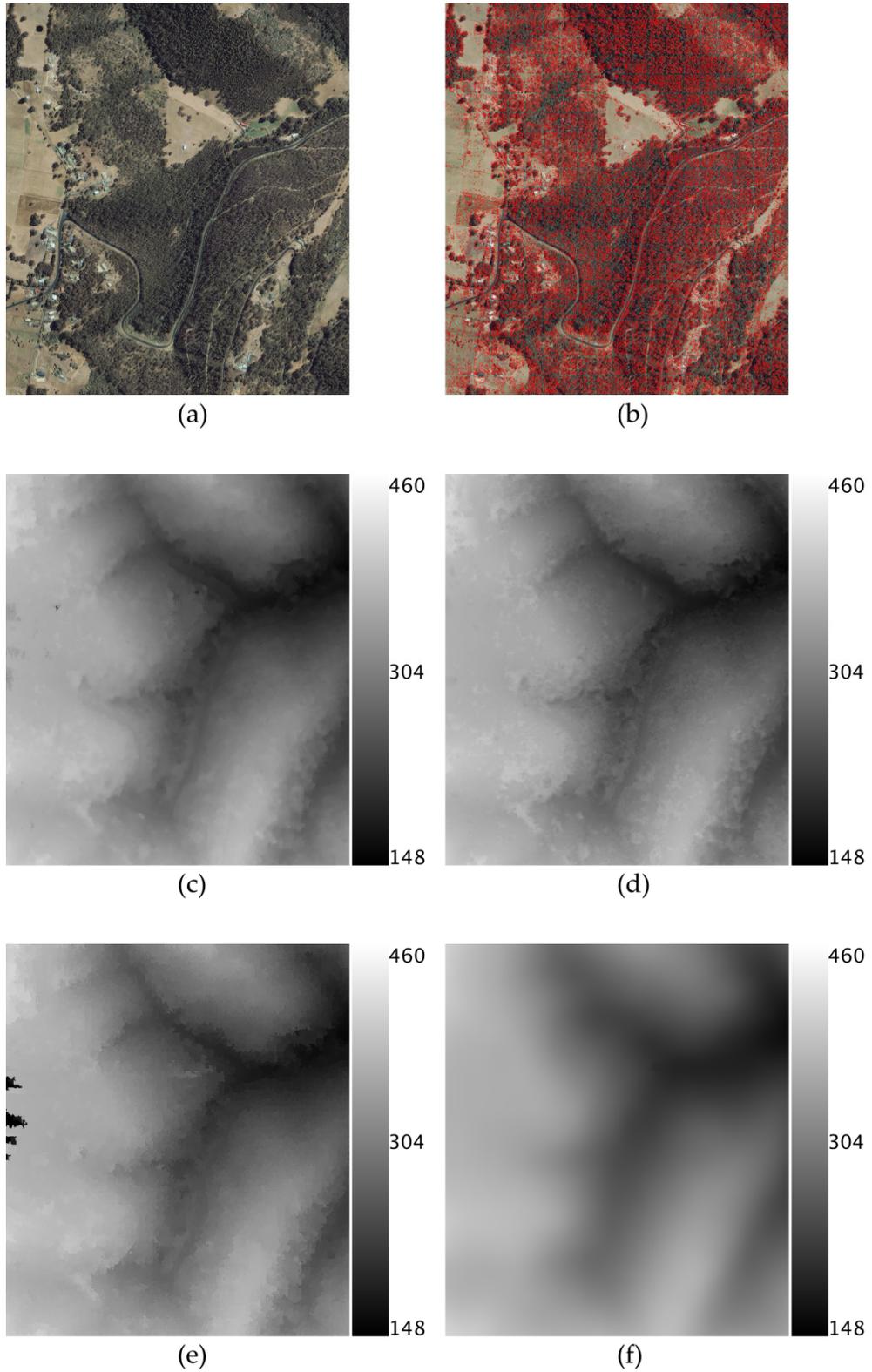


Figure 7.36. Interpolation results for various algorithms on patch 6 (1189x1369): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

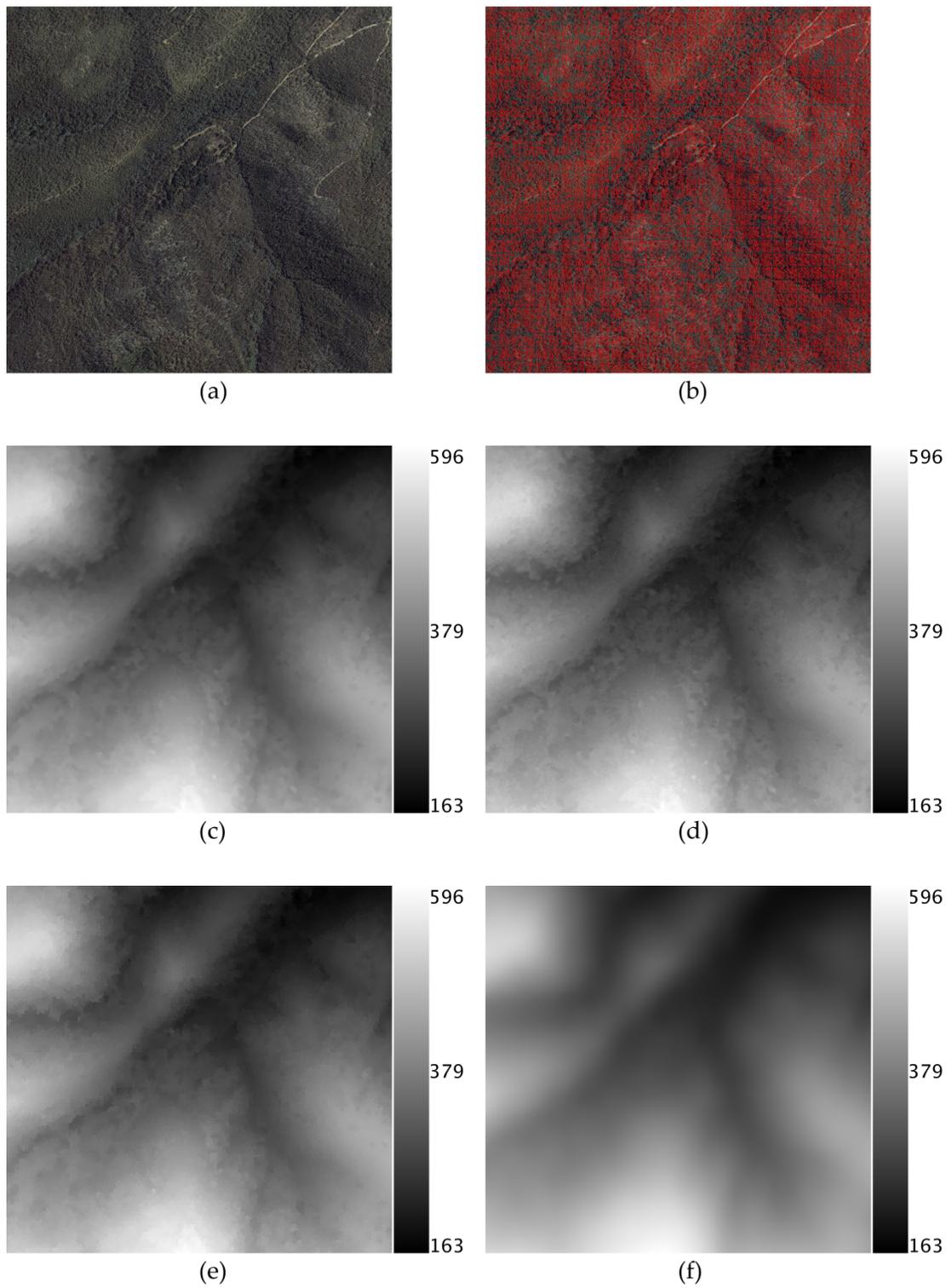


Figure 7.37. Interpolation results for various algorithms on patch 7 (1868x1791): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

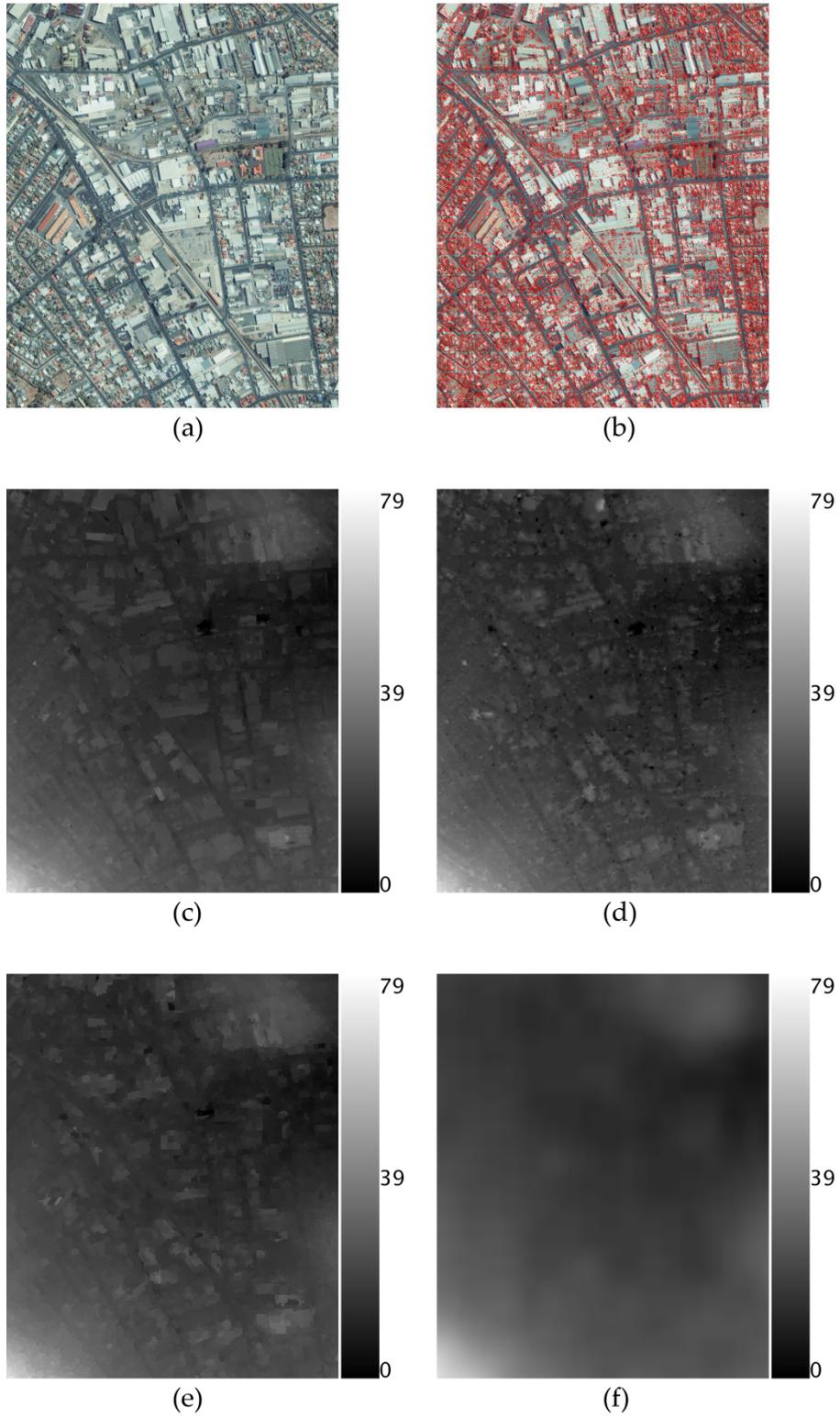


Figure 7.38. Interpolation results for various algorithms on patch 8 (1121x1378): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

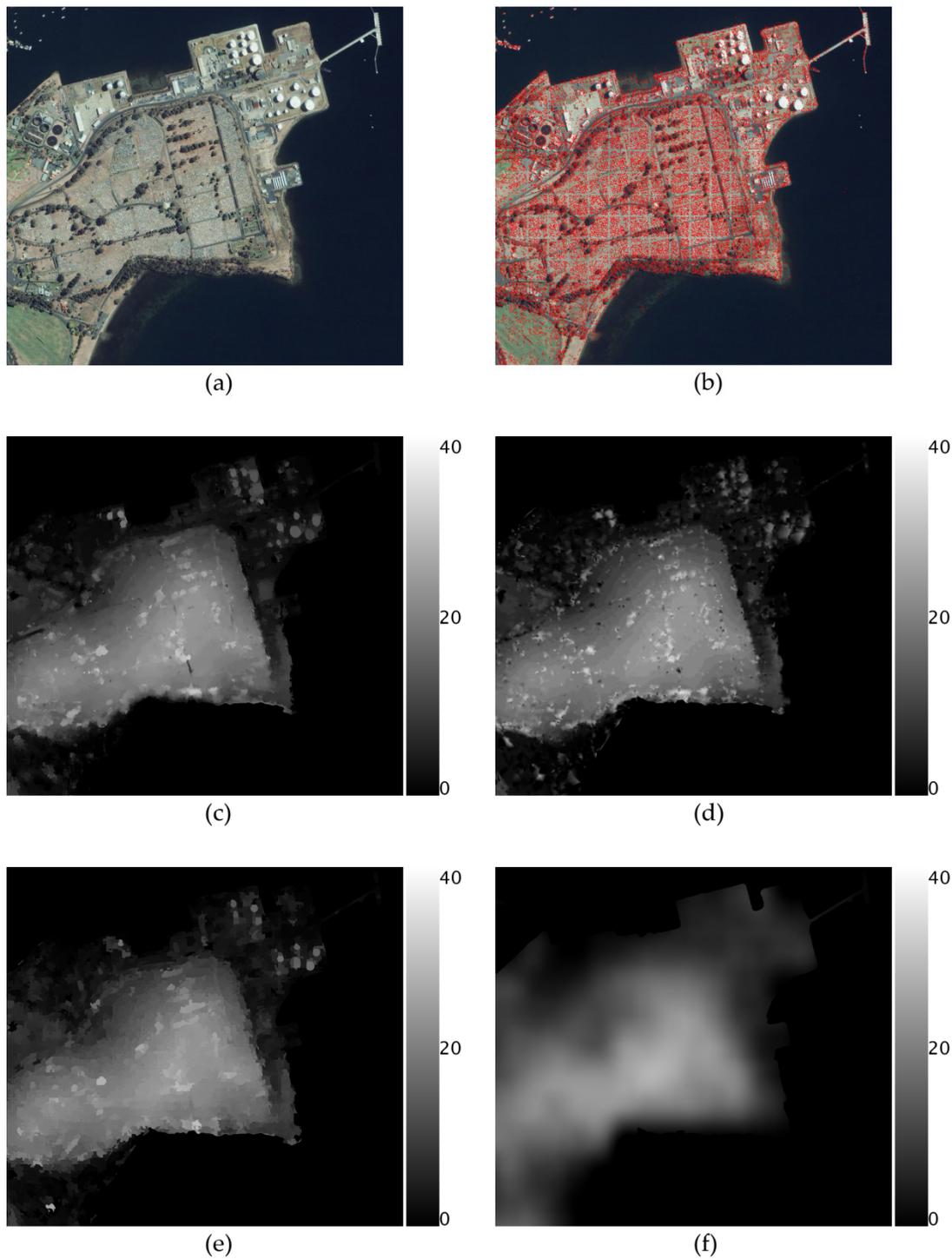


Figure 7.39. Interpolation results for various algorithms on patch 9 (1185x1079): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

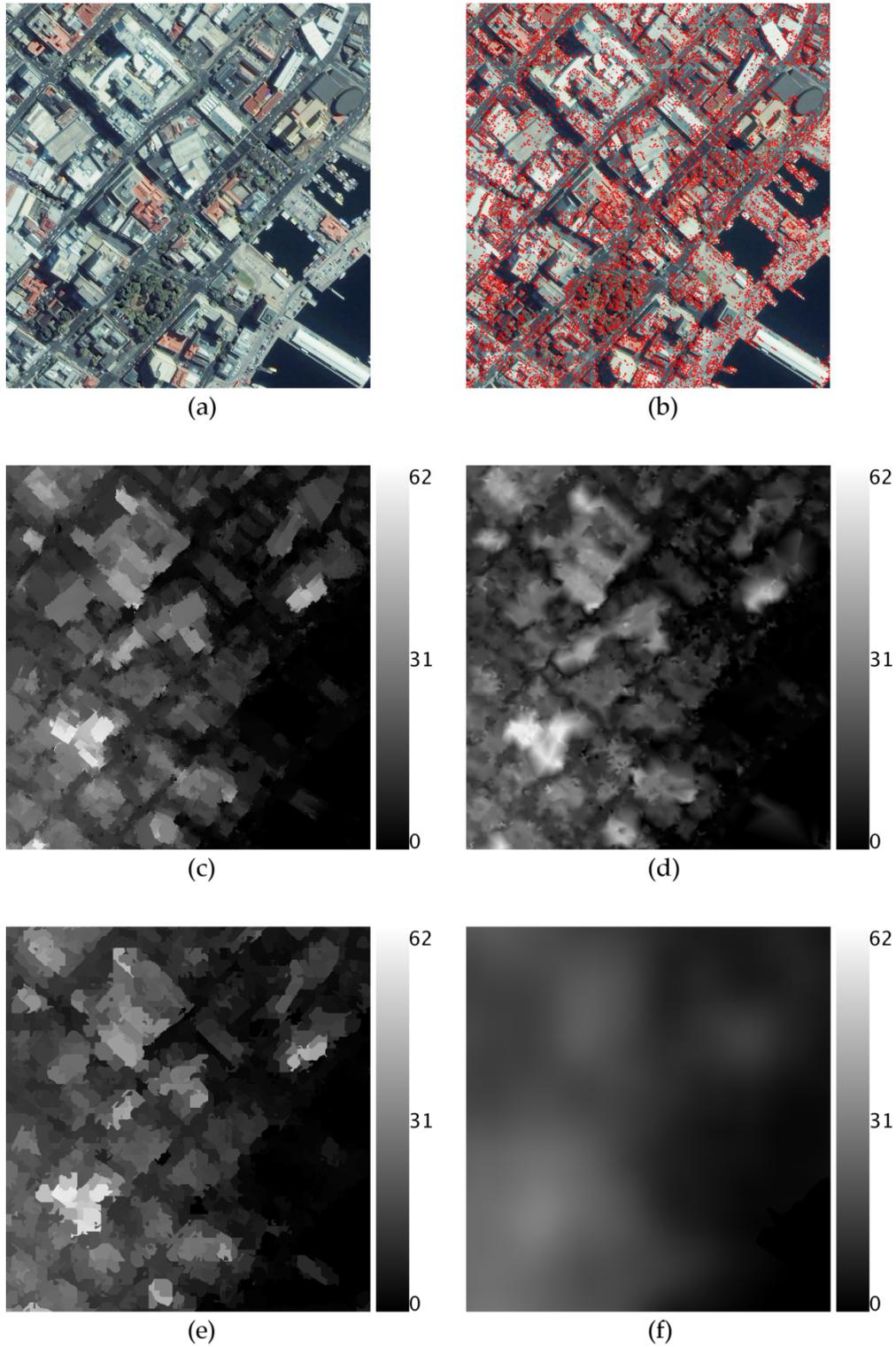


Figure 7.40. Interpolation results for various algorithms on patch 10 (581x619): (a) original image, (b) feature points, (c) Edge aware filter with 20 iterations, (d) Meshing, (e) Segmentation, and (f) SRTM registration

The results for interpolation by integral images (rectangular windows) are as follows:

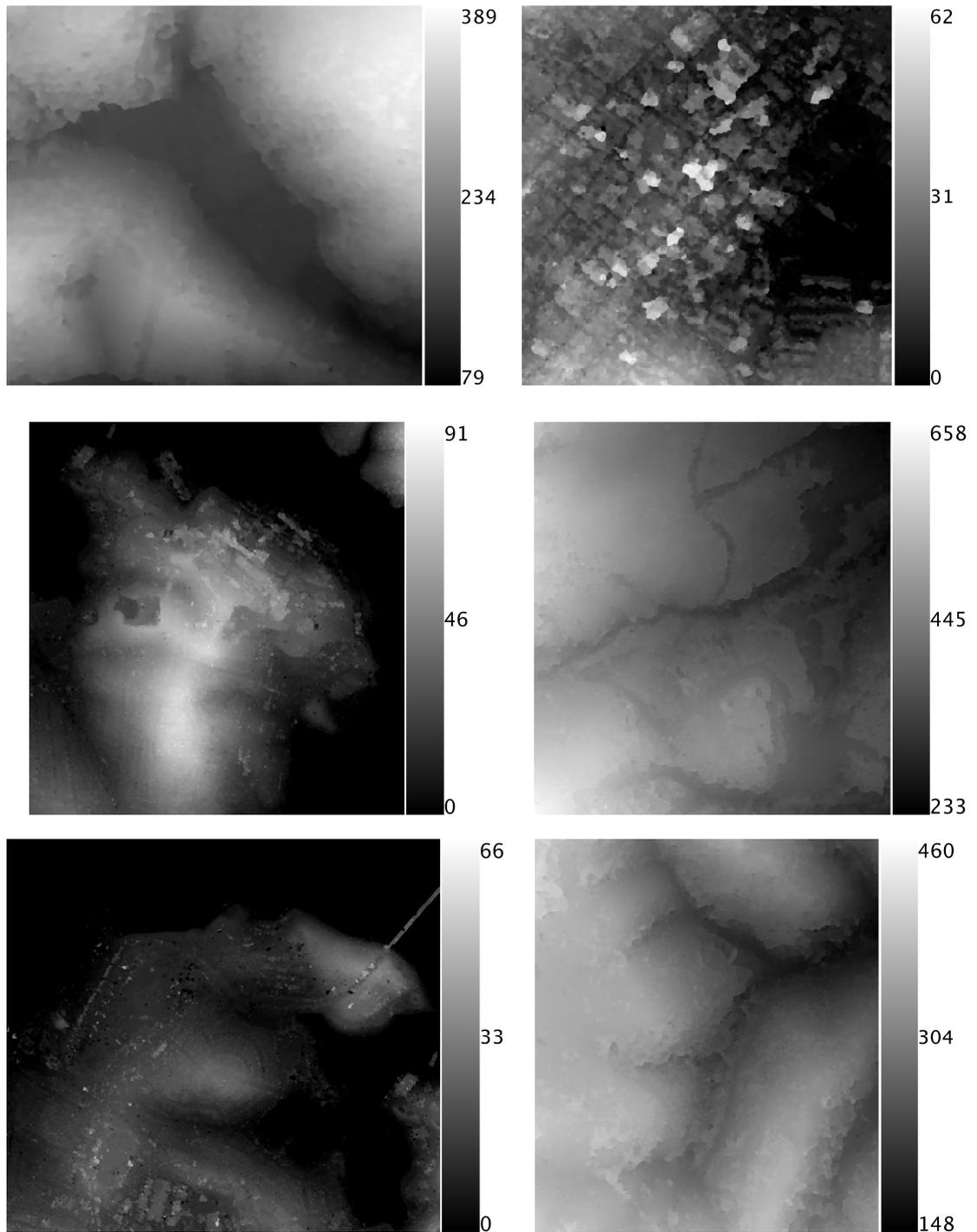


Figure 7.41. Results for interpolation by integral images (rectangular window).

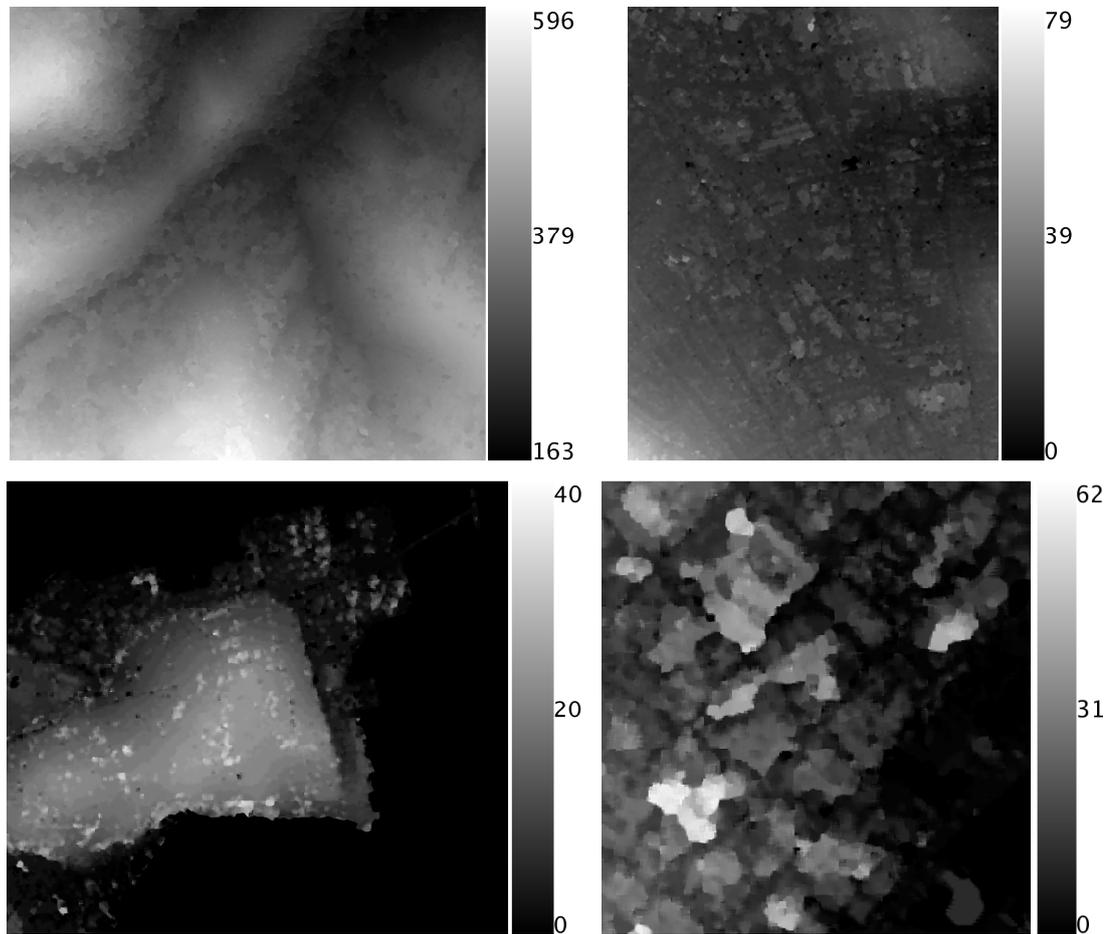


Figure 7.41. (cont'd.) Results for interpolation by integral images (rectangular window).

The results for interpolation by integral images (circular windows) are as follows:

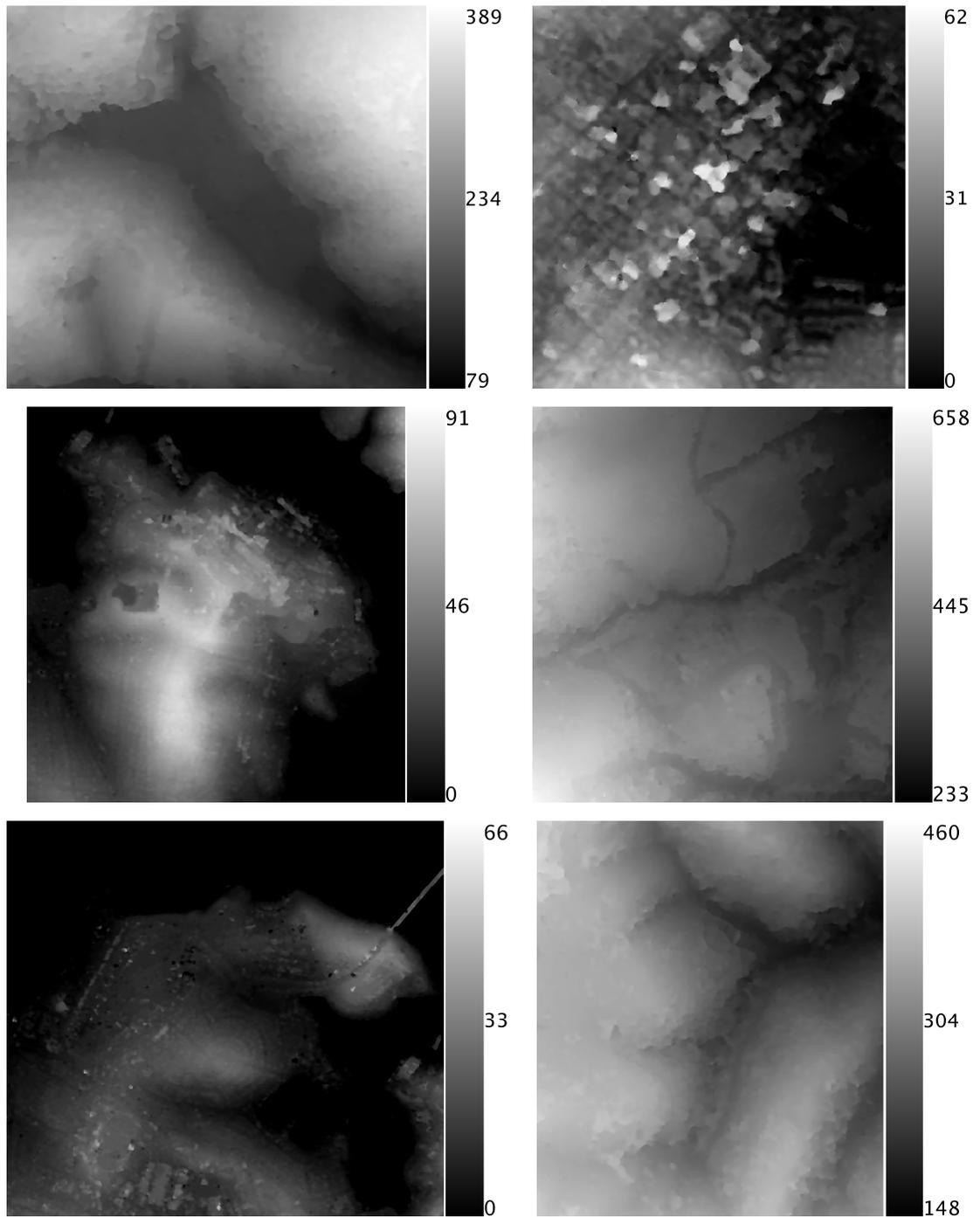


Figure 7.42 The results for integral image based interpolation using circular windows with a radius of 10 pixels.

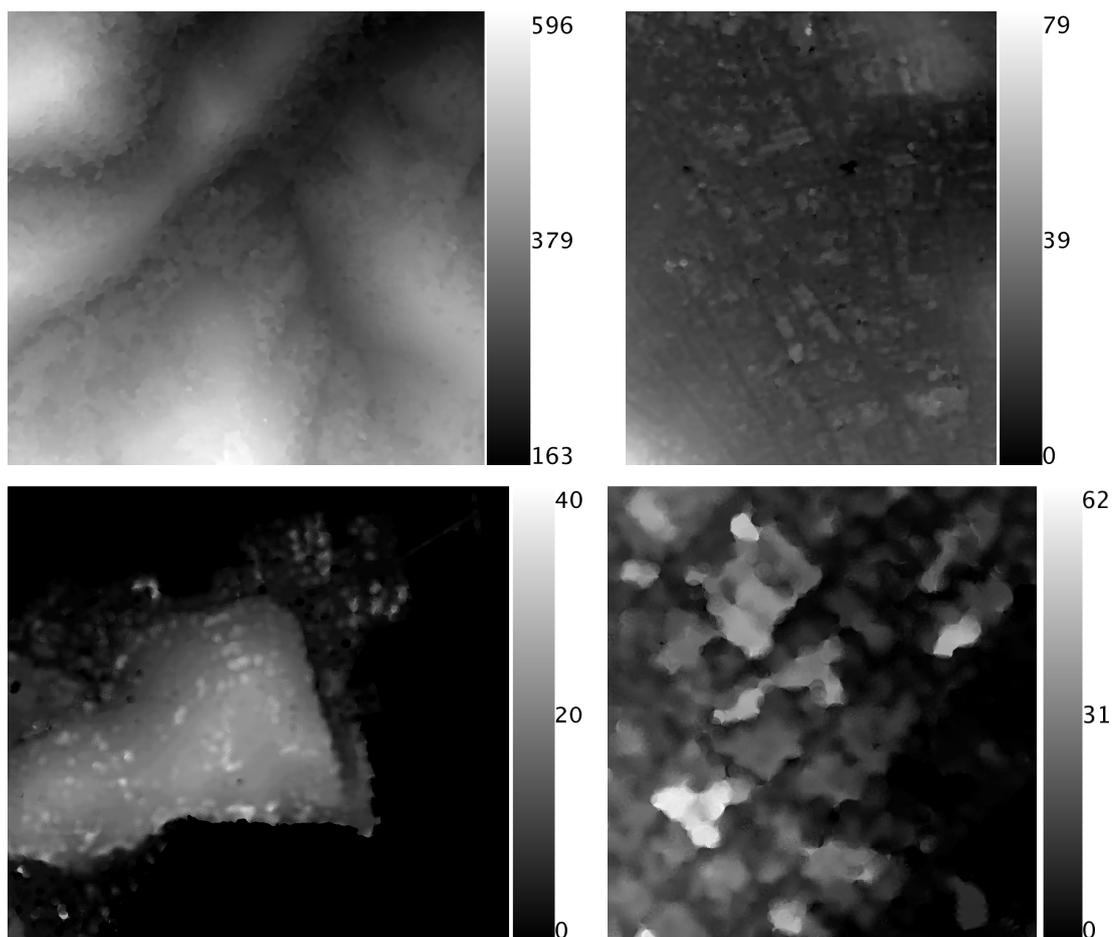


Figure 7.42 (cont'd.) The results for integral image based interpolation using circular windows with a radius of 10 pixels.

The experiments on the algorithms demonstrated that, the algorithms had advantages and disadvantages in different regions. The results can be assessed as follows:

- i) Edge-aware filtering is superior for dense urban zones, where there exist sharp object boundaries and the density of the sparse correspondences is relatively low. The algorithm can compensate for the insufficient number of correspondences through interpolations and a sharp DSM is generated. On the other hand, in the forests, where

there exist trees with similar colors the algorithm oversmooths the region. It should also be noted that, the performance of the edge-aware filtering algorithm is higher for color images.

- ii) Meshing is superior in forests, where the density of the sparse correspondences is large. The algorithm is not dependent on the image color characteristics, thus no blurring occurs in forests. However, the performance in dense urban zones is lower due to the insufficient number of correspondences.
- iii) SuTP-based approach is dependent on the density of the sparse correspondences. The algorithm's performance is similar to meshing in the forests, and between meshing and the edge aware-filtering in the urban zones. One advantage of this algorithm is that, the regions where the number of correspondences is insufficient can easily be determined. Obviously, this is also possible with meshing (by checking the areas of the triangles). Still, SuTP is more informative since it utilizes the color/intensity information of the image.
- iv) Interpolation with the integral image approaches cannot generate sharp DSMs. Rather, they can generate smoothed DSMs. Although this can cause loss of detail, the smoothing behavior reduces the artifacts of false matches. The smoothness can be controlled easily by changing the window radius.

7.6.2 Comparison with COTS Software

Interpolation was also performed with a commercial software package, namely PCI Geomatica. PCI Geomatica has the capability of RFM reconstruction, bundle adjustment and dense DSM generation.

The software was run on the IKONOS Hobart dataset, which includes 123 GCPs that are visible in both images. After bundle adjustment with 123 GCPs, sub-pixel projection errors were obtained in both images.

For the DEM generation preferences, the software was asked to fill the gaps that might be caused by various reasons such as occlusion zones and failures in matching. For comparison, the edge-aware filter interpolation algorithm is selected. The results are presented in the figures below.

The results of the proposed method and PCI Geomatica are not aligned: PCI Geomatica warps the satellite images to obtain “epipolar images” and outputs the resultant DSM for the warped reference image, while the proposed scheme generates the DSM for the original reference image. It must be noted that, the results below are the height images which are not georeferenced. Thus, PCI Geomatica or the proposed method cannot be compared for geolocation accuracy by investigating the following figures. One should expect better geolocation accuracy in the PCI Geomatica output since the software uses block adjusted RFMs obtained by utilizing GCPs.

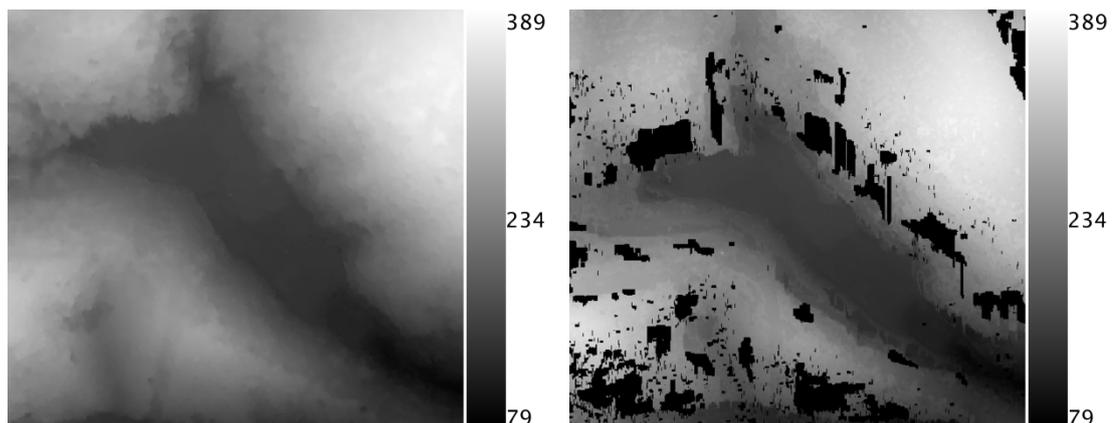


Figure 7.43. Comparison with PCI Geomatica, patch 1 (a flat region in a forest); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

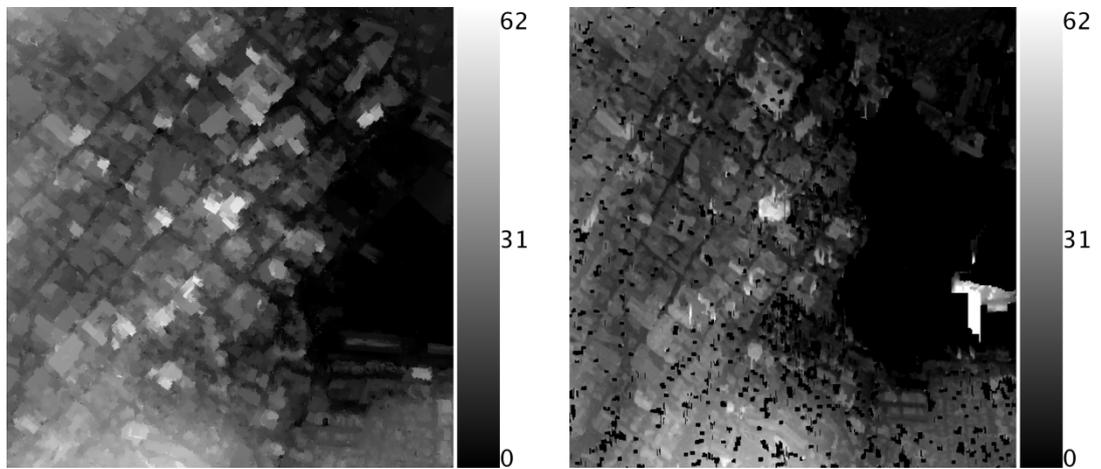


Figure 7.44. Comparison with PCI Geomatica, patch 2 (urban); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

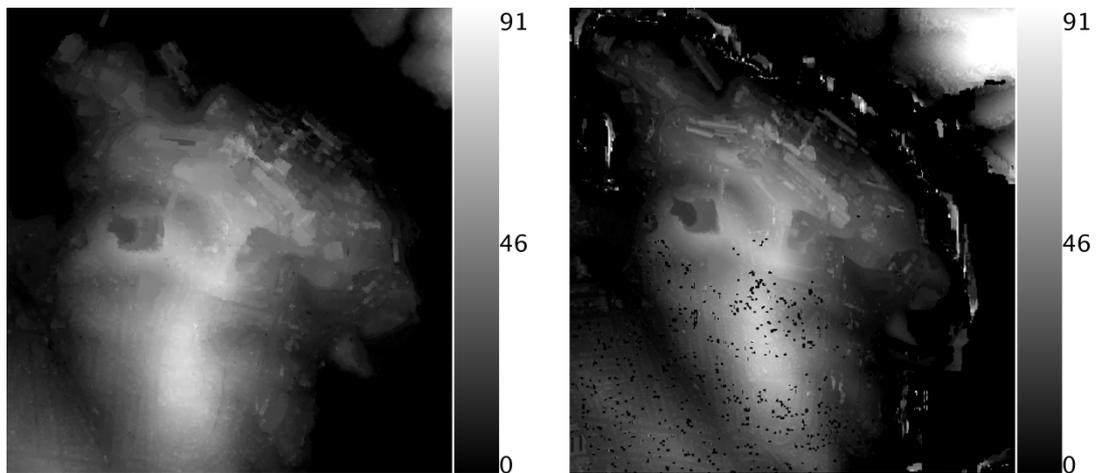


Figure 7.45. Comparison with PCI Geomatica, patch 3 (industrial); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

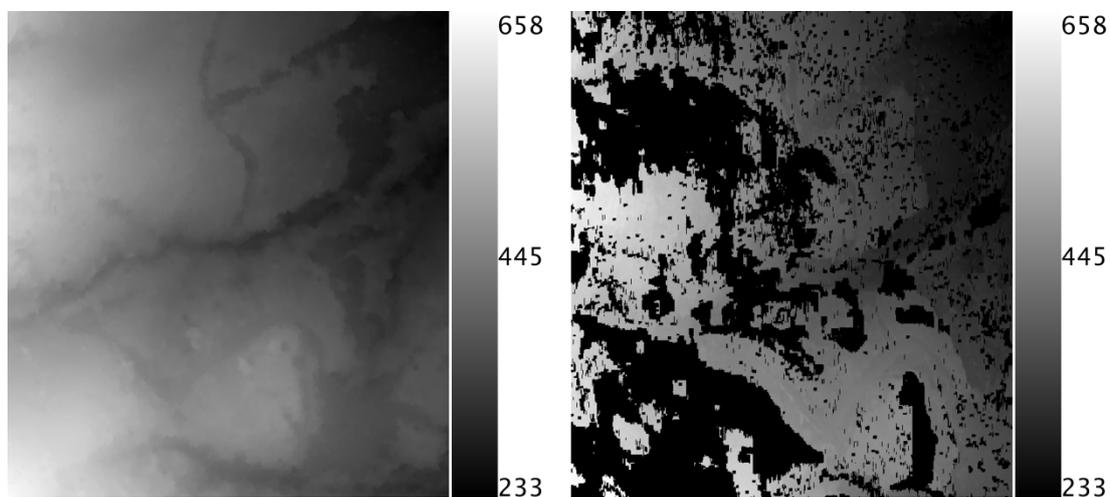


Figure 7.46. Comparison with PCI Geomatica, patch 4 (forest); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

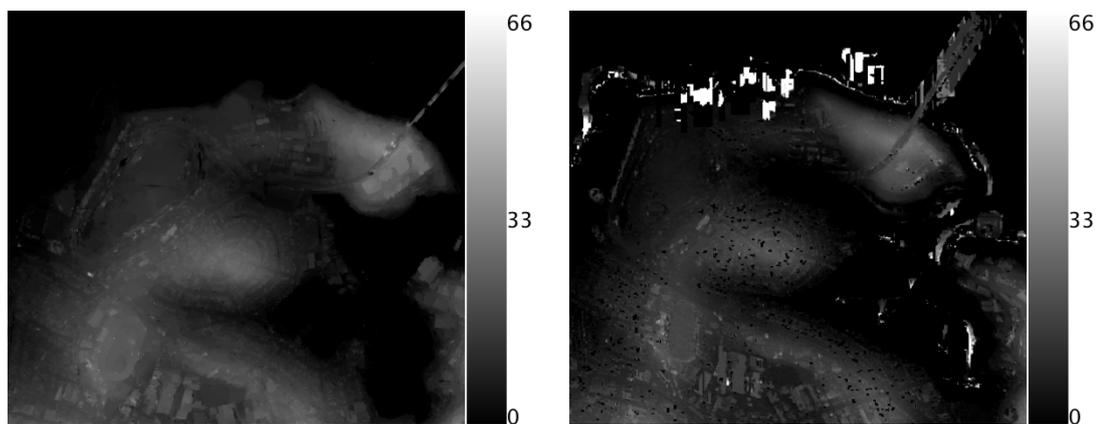


Figure 7.47. Comparison with PCI Geomatica, patch 5 (industrial and suburban); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

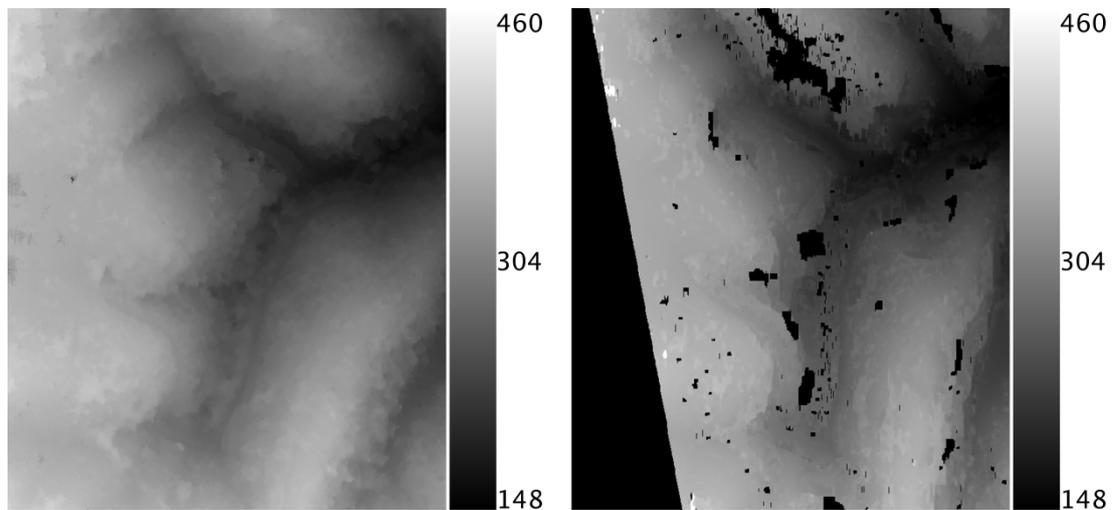


Figure 7.48. Comparison with PCI Geomatica, patch 6 (forest); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

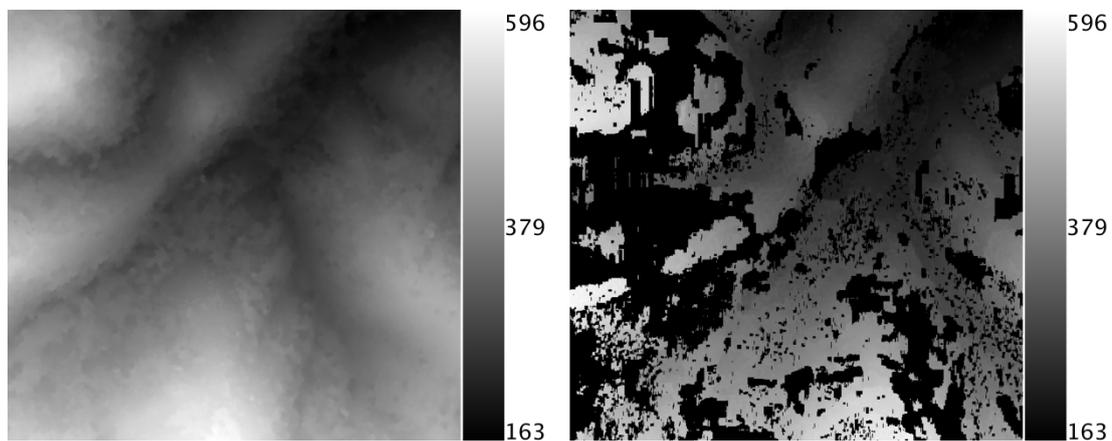


Figure 7.49. Comparison with PCI Geomatica, patch 7 (forest); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

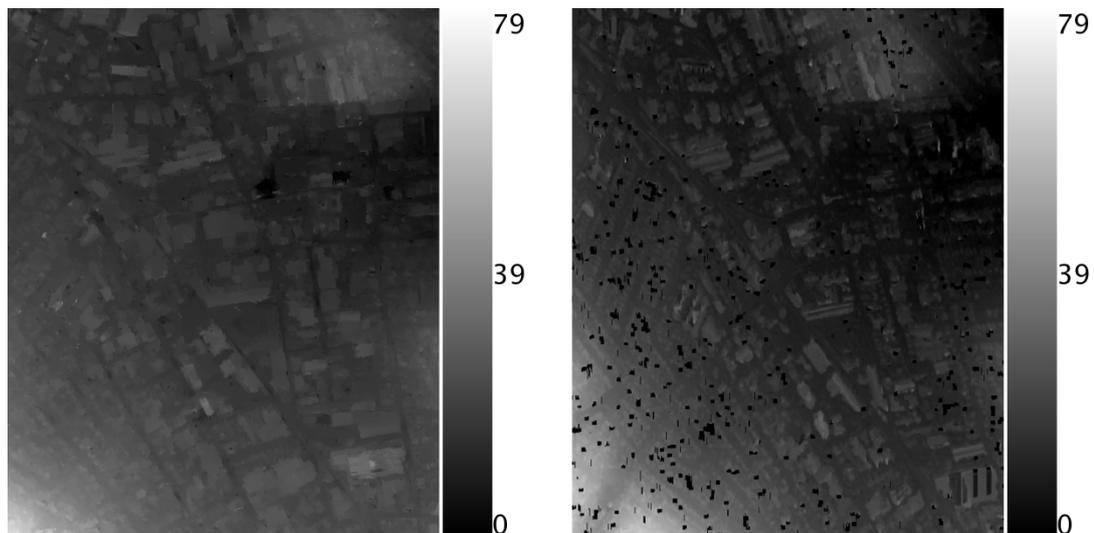


Figure 7.50. Comparison with PCI Geomatica, patch 8 (industrial); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

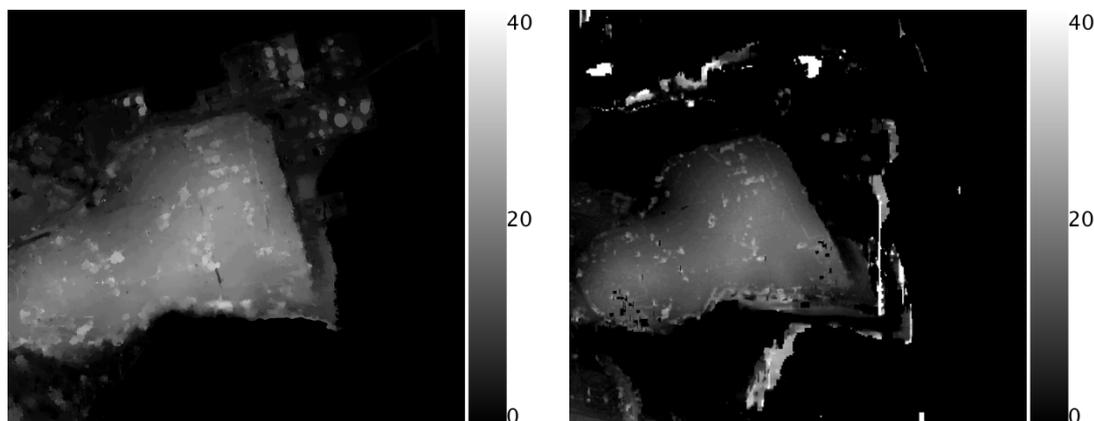


Figure 7.51. Comparison with PCI Geomatica, patch 9 (industrial); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

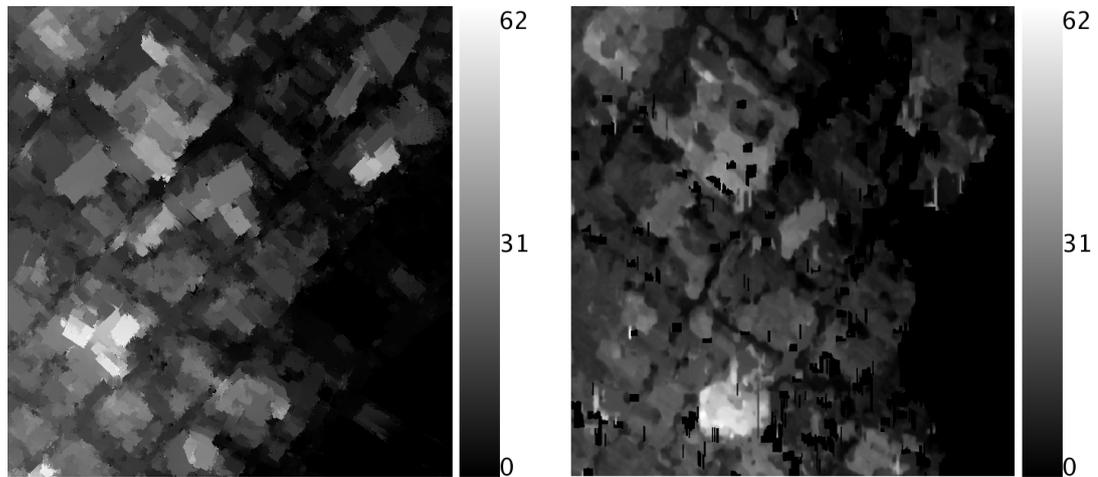


Figure 7.52. Comparison with PCI Geomatica, patch 10 (urban); edge-aware filter (left) and PCI Geomatica (right) outputs. The colorbars are in meters.

The results demonstrate that the proposed DSM generation scheme provides comparable performance with the COTS software that utilizes the GCPs. The performance of the proposed methodology is clearly superior in the forests where PCI Geomatica output has large gaps. PCI was observed to provide slightly better results in the suburban and industrial regions, although there appear small gaps in the output. In the urban zones, the outputs of the proposed scheme and PCI Geomatica are comparable: PCI geomatica output has less noise and the output of the proposed scheme provides more detail. In all zones, the gaps of the PCI Geomatica output are required to be edited manually; except for the forest zones, where the gaps are too large for proper results. The blunders in the shores are also to be corrected manually or by segmenting the pixels in the sea.

7.7 Execution Times

The timing of the methods was measured on Ubuntu Linux running on a PC with Quad-Core Pentium Q9550 Processor, 8GB DDR2 1300 MHz RAM. All tasks were run using a single core of the processor, but the quad-core architecture allowed dedicating a single core completely on the processing tasks. The processing times are presented in the following sections.

7.7.1 SRTM Registration

SRTM Registration timing results are presented in (Table 7.5).

Table 7.5. Processing times for SRTM registration

Dataset	Image Size (pixels)		Processing Time (s)
	Width	Height	
IKONOS Hobart	12122	13148	81
IKONOS Eskişehir	6968	5728	47
IKONOS San Diego	2001	2001	37
WorldView-1 La Mola	10000	10000	88
WorldView-1 Terrassa	10000	10000	89
WorldView-1 Vacarisses	10000	10000	91
CARTOSAT-1 La Mola	2066	2784	80
CARTOSAT-1 Terrassa	2488	2782	77
CARTOSAT-1 Vacarisses	2488	2782	77
ALOS Norikura	4992	16000	141
ALOS Hiroshima	4992	16000	130

7.7.2 Sparse Reconstruction

Feature Detection

In this study, matching with FAST features is performed by using the BRIEF descriptor and to avoid redundancy; for each feature point, the descriptor is computed and stored only once. Therefore, BRIEF descriptor computation time is included in the FAST feature detection lines of the tables below.

Table 7.6. Feature detection times (seconds) for IKONOS images

		IKONOS Hobart	IKONOS Eskişehir	IKONOS San Diego
Image Size	Height	13148	5728	2001
	Width	12122	6968	2001
Harris		11	4	1
FAST+BRIEF		156	17	3
GFTT		11	4	1

Feature detection algorithms can generate insufficient number of interest points for the WorldView-1 and CARTOSAT images and are accepted to fail. Thus, timings for these datasets are not given here.

Table 7.7. Feature detection times (seconds) for ALOS images

		ALOS Norikura	ALOS Hiroshima
Image Size	Height	16000	16000
	Width	4992	4992
Harris		8	6
FAST+BRIEF		9	22
GFTT		8	6

As seen in the tables, Harris and GFTT detectors have the same values. This is because the Harris features are obtained by eliminating the GFTT points with a higher threshold for the smaller eigenvalue of the structure tensor.

Matching

The timings of sparse matching is presented in Table 7.8 and Table 7.9. The “Template Match” rows present the execution times for pyramidal template matching and the “BRIEF” rows state the execution times for the BRIEF matching step of the BRIEF+TM algorithm. As it can be seen in the tables, for the BRIEF+TM algorithm, the main computational load is in the BRIEF step (which is faster than pyramidal template matching) and the refinement step (Template Matching) corrects the match points in a short time. However, when the density of the feature points is high (e.g., large forest regions in the Hobart scene), the number of candidate BRIEF descriptors in the search region is also large and many score computations are required. In such a case, the benefits of utilizing the BRIEF descriptor is questionable (regarding the execution time) since descriptor computation is also required (in the feature detection phase) for BRIEF.

It must be noted that, the used OpenCV implementation of the KLT algorithm utilizes all CPU cores, thus the actual complexity of the algorithm is approximately four times larger.

Table 7.8. Matching times (seconds) for IKONOS images

	IKONOS Hobart		IKONOS Eskişehir		IKONOS San Diego	
	Num. of features	Time	Num. of features	Time	Num. of features	Time
KLT	6,329,828	197	1,200,701	45	120,227	6
BRIEF	7,784,124	227	769,182	65	117,873	8
Template Match	7,784,124	585	769,182	91	117,873	12
BRIEF+TM	7,784,124	429	769,182	88	117,873	9

Table 7.9. Matching times (seconds) for ALOS images

	ALOS Norikura		ALOS Hiroshima	
	Num. of features	Time	Num. of features	Time
KLT	919,696	72	3,685,098	190
BRIEF	Failed		1,032,197	59
Template Match	Failed		1,032,197	105
BRIEF+TM	Failed		1,032,197	90

In all cases, the sparse matching step could be performed in short time for all algorithms when the proposed initial estimate generation scheme was used.

Reconstruction

Reconstruction was performed on the satellite images for which the matching was successful. The second column in the following table presents the number of processed correspondences, which determines the processing time. As mentioned earlier, in the reconstruction phase, the correspondences

with high re-projection errors were eliminated. The resulting sparse points were approximately 10% less in number. As seen in the table, the reconstruction time is proportional to the number of processed correspondences.

The processing times are given either for BRIEF-TM or GFTT-KLT pairs. Hobart points were obtained from the BRIEF-TM algorithm, The Eskişehir points were obtained by combining the outputs of the two approaches. The remaining timings are for the GFTT-KLT algorithm outputs. The selection of the algorithm was based on the number of correspondences generated, except for the Norikura case, where only GFTT-KLT could provide successful correspondences.

Table 7.10. Reconstruction times (seconds) for different datasets

Dataset	Number of points	Processing Time (s)
IKONOS Hobart	7,030,256	210
IKONOS Eskişehir	1,512,055	30
IKONOS San Diego	120,227	4
ALOS Norikura	919,696	30
ALOS Hiroshima	1,417,393	44

The difference between the number of points in Table 7.8, Table 7.9 and Table 7.10 is caused by the outlier elimination step, which is performed before the reconstruction phase. In Table 7.8 and Table 7.9, the number of processed feature points is given, since the execution time depends on the number of feature points, regardless of the number of successful matches.

7.7.3 Dense DSM Generation

In this section, interpolation times for the processed datasets are given. Again, the WorldView-1 and CARTOSAT-1 datasets are excluded due to the failure in matching.

The interpolation radius of the integral image approaches were selected according to the density of the sparse points. The radius (R) was 10 in all datasets, except for the Norikura image (R=50), for which, the density of the correspondences was low. The interpolation window size was not effective for rectangular interpolation, but in the circular interpolation case, the processing time increased linearly with the interpolation radius.

Table 7.11. Interpolation times (seconds) for IKONOS images

		IKONOS Hobart	IKONOS Eskişehir	IKONOS San Diego
Image Size	Height	13148	5728	2001
	Width	12122	6968	2001
Edge-Aware Filter		230	61	6
Int. Img. Rect.		56	15	0
Int. Img. Circ.		88	30	1
Segmentation		100	33	3
Meshing		23	5	2

For the ALOS Norikura images, insufficient number of correspondences was obtained due to poor image contrast and radiometric inequalities between the image columns. Therefore, for the integral image approaches, the number of iterations (passes) to fill all the pixels of the image was large. The interpolation radius was determined using the number of pixels in the

image and the number of sparse reconstruction points. As a rule of thumb, the interpolation radius can be selected as: $R = \sqrt{N/P}$, where R is the window radius, P is the number of sparse points and N is the number of pixels in the region of interest. If there exists large regions where DSM is not expected (e.g. sea), the N must be determined accordingly.

Table 7.12. Interpolation times (seconds) for ALOS images

		ALOS Norikura	ALOS Hiroshima
Image Size	Height	16000	16000
	Width	4992	4992
Edge-Aware Filter		124	126
Int. Img. Rect.		30	34
Int. Img. Circ.		617	121
Segmentation		58	89
Meshing		7	7

The overall execution time for dense DSM generation with the COTS software depends on the image resolution and the execution times are not reported in the photogrammetry literature. However, it is known that the entire processing workflow may take hours or days, depending on the image sizes and the computer hardware.

The proposed scheme was compared with COTS software on IKONOS dataset for the execution time. The selected software was PCI Geomatica, which has the capability of bundle adjustment, RFM reconstruction, and dense DSM generation. The DSM generation on PCI Geomatica took 6 hours and 8 minutes of execution time. Sample patches of the DSM that were generated by PCI Geomatica are presented in Section 7.6

The overall execution time in this study is in the order of minutes, depending on the algorithm choice in each step. When the slowest algorithms (Harris + SRTM registration + Template Matching + Sparse reconstruction + Edge-aware filter) were selected on the largest-overlap colored stereo data (IKONOS Hobart), the processing time became 17 minutes and 16 seconds. The execution time reduced to 7 minutes and 47 seconds on the same data, when the fastest algorithms were selected.

CHAPTER 8

CONCLUSION AND DISCUSSION

8.1 The Significance of the Study

In this study, a complete, GCP-free DSM generation scheme is developed and implemented for the first time. The results on various satellite images demonstrated that a flexible and accurate DSM generation scheme is practical in the absence of GCPs.

The developed scheme is driven by the motivation to improve the knowledge on the projection model with different sources of information and the idea behind the solution brings a new way of thinking in various phases of the DSM generation workflow.

The first contribution is in the registration of LR-DSMs to the satellite images. The current state-of-the-art in the literature uses precise RFMs obtained by using GCPs to generate a sparse point cloud and tries to register this point cloud with the available LR-DSM. Such a scheme is impractical in smooth regions and the registration error may be high, since the registration is generally done with correlation, which is unreliable in smooth regions due to the existence of noise in the sparse point cloud and the LR-DSM. With the proposed approach, no sparse reconstruction is required and the performance in the smooth regions is solid. This scheme also eliminates the

requirement for stereo images for registration as the need for sparse 3-D reconstruction is eliminated. Each satellite image can be registered with its projection functions simply by projecting the LR-DSM onto the image domain.

The sparse correspondence generation problem is also re-visited and registration of the LR-DSM is utilized for initial estimate generation in the sparse reconstruction problem. Classically, in the photogrammetry literature, the search is reduced to 1-D using the epipolar curves and the search region is further narrowed down using a coarse DSM obtained from the higher levels of the image pyramid. In the absence of GCPs, the RFMs are not accurate enough to reduce the correspondence generation problem down to a 1-D search and the image domain errors in both image1 and image 2 result in a wide “epipolar strip”, in which the search is expensive. Thus, the epipolar curves cannot be used to improve the computational efficiency. As an alternative, in this study, it is shown that the SRTM registration can provide very accurate initial estimates for the match points and fast sparse correspondence generation can be performed using the recent, descriptor-based computer vision techniques or optical flow. It is also shown experimentally that many accurate correspondences can be generated in short time using a hybrid method (BRIEF + Template Matching) or optical flow (pyramidal KLT).

This approach differs significantly from the current trend in photogrammetry, in which the matching is performed with an expensive mature method (ILS). The ILS performs affine transformation estimation and interpolation-based warping for each stereo correspondence. However, the results obtained in this study demonstrate that, the transformation between the patches around the match points are simple translations thanks to the

advanced camera control techniques used in the imaging process of modern, high resolution, narrow-FOV EOSs. As the resolution increases, the validity of the assumption that the camera motion vector is constant for the adjacent image rows improves and a template that is sufficient for matching is less affected by the orbital motion of the satellite. As a result, in many cases, an effort for affine parameter estimation is not required in correspondence generation.

The proposed correspondence generation strategy is very successful in the forests, where the methods utilized in the photogrammetry literature are known to be problematic. Contrary to these methods, the proposed method can generate many accurate correspondences in the forests. The density of the correspondences is high enough to use meshing for obtaining a correct DSM.

In this study, dense DSM is obtained through various interpolation schemes including meshing, segmentation, averaging (with integral images) and edge-aware filtering. These approaches are shown to provide successful results for various types of textures (forests, urban and rural regions, etc.) in the satellite images.

The photogrammetry literature on the interpolation phase is shallow and the methods are limited to meshing and region growing. Meshing is used in the intermediate steps for generating a coarse DSM and region growing is a quite expensive technique that performs matching for all points in the image. In this study, it is shown that, such and expensive method is not required provided that the number of sparse correspondences is sufficient.

Another contribution of the study is an RFM bias equalization scheme that uses the registered LR-DSM and accurate stereo correspondences. It is

shown that, with bias equalization, the reconstructed height accuracy can be improved from 4 m to 1.7m for IKONOS images.

It is also shown that, in the existence of LR-DSM registration, the parameters of simple empirical imaging models can be computed thanks to the availability of ground information for every pixel in the image. The models obtained with this method provide projection accuracies better than 3 pixels (rms) when they are computed for chunks of image rows. With this observation, the concept of piecewise empirical models is proposed. Such models are similar to the rigorous models, which are based on the ephemeris, whose sampling period is much larger than the acquisition time of a single row.

8.2 Usage

The DSM generated with a GCP-free workflow can be used in appropriate applications, in which the accuracy is not the major concern or the GCPs cannot be collected.

EOSs work continuously and record a great amount of imagery per unit time. The GCP collection process cannot be performed with the same speed, thus a GCP-free scheme is valuable.

There may be many applications that do not require sub-pixel accurate DSMs, but the additional information on the DSM (other than LR-DSMs) may be useful. One such application is the Google Earth. Google Earth uses LR-DSMs and few manually generated building models for height information. With the proposed DSM generation scheme, even the short buildings can be differentiated. Although the generated DSM is somewhat noisy in the urban zones, with a large map database such as Google Maps,

these errors can be corrected. For a very-large-coverage project like the Google Earth, collection of GCPs is not feasible, except for the big cities in the developed countries, whose LIDAR data may be available. Thus, a GCP-free workflow is essential for such large scale projects.

In this study, various interpolation schemes are presented. It is observed that different methods are successful in different regions. For example; the edge-aware filter generates good results in the regions where the correspondences are sparsely distributed (due to occlusions, etc.) and meshing generates poor results in these areas since it does not use visual clues from the satellite image. On the other hand, in the forests, where the trees look alike, the edge-aware filter pays a high price for trusting in the color similarity and generates an “oil-painted” DSM. In forests, the blindness of meshing to the visual clues is awarded and a sharp DSM is obtained thanks to the very high density of the correspondences. One advantage of these methods is that, both are very fast. Thus, one can run both algorithms on the same stereo image pair and take the result of the better algorithm for each region.

8.3 Remaining Problems

SRTM is used as the LR-DSM. Since SRTM is a DSM, wide and tall buildings are observable as smoothed hills in the SRTM data. Thus for such regions, the matching scheme proves successful. On the other hand, the matching algorithms assume Lambertian surfaces, which have uniform reflection directivity. If that is not the case, there may be a significant radiometric difference in the stereo images, and the matching algorithms fail to find the match points or generate false correspondences. In fact, this

problem is not specific to tall buildings; when there is a significant change in the observed gray levels, the matching processes encounter difficulties.

The proposed interpolation schemes use the sparse reconstruction results as the seed points in the interpolation. It is observed that, such interpolation schemes are dependent on the number and the distribution of the sparse seed points, which are weak in the occlusion zones caused by tall buildings or steep terrain. In these regions, the algorithms fail to detect visible matches or generate false correspondences. Although many outliers can be eliminated using neighbor consistency, the results are not error-free. This is a typical problem in stereo correspondence generation and various approaches have been proposed in both photogrammetry and computer vision literature. However, the majority of such approaches utilize the epipolar curves/lines, which are unavailable in a GCP-free study.

Another typical problem in the urban zones is the false matches caused by moving objects such as cars. Detection of moving objects is difficult due to the large B/H ratio in the stereo imaging scenario. This problem remains unresolved.

Large water bodies also cause problems in the matching process due to the moving waves. These regions can easily be detected in colored images, but the problem remains for gray level images. This problem can be solved by using external information on the water bodies, such as geographic maps. Even very low resolution maps will suffice for this task since the reconstruction will be good for shorelines.

8.4 Future Work

The problems in the occlusion zones can be handled by using triple images. Such effort requires appropriate match generation and reconstruction scheme. The points must be categorized by the satellite images that they are observable. Due to the large number of correspondences, this must be performed automatically, which means automatic detection of the occlusion zones. The points that can be observed in the three images should be subjected to a 3-image reconstruction algorithm with the appropriate cost function on 3 images and a 6x3 Jacobian matrix.

The problems observed in the occlusion zones must be addressed carefully. This problem is serious in a GCP-free workflow, since the occlusions cannot be detected using epipolar curves. However, the density of the correspondences reduces significantly, where feature matching has the potential to fail. With this clue, it is possible to reconsider the correspondences obtained in the “few-correspondence zones”. In these regions, various approaches can be tested for better matching performance. This re-work shall be performed after a previous matching step which handles the great majority of the correspondences. This way, more complicated methods can be run to match the remaining feature points, which are much less in number.

Although the KLT algorithm performs sub-pixel matching, the decision for accepting a correspondence is based on the squared error, which is dependent on illumination. Thus, the correlation coefficient is more frequently used in the photogrammetry literature due to its illumination invariance. The method presented in this study does not perform sub-pixel

matching with the correlation coefficient. This can be achieved by fitting a polynomial to the correlation coefficient result in the search window, around the peak value. The local maxima of this polynomial can be used as the sub-pixel match location. Still, one must be aware of the fact that, in a GCP-free study, the benefits of sub-pixel matching is questionable since the projection functions are not perfect anyway.

The processing time of the algorithms presented here can be reduced further with parallel processing. Almost all of the processing phases are parallel in nature and thus efficient parallelization is possible.

The approaches presented here can also be used in the studies when GCPs are available. Of course, bias equalization is not applicable to such studies, since the usage of GCPs provides much better projection accuracy. Still, LR-DSM registration, correspondence generation and interpolation steps are applicable when GCPs are available. The existence of GCPs result in very accurate projection models enabling the usage of epipolar curves. In that case, matching along the epipolar curve will probably be more efficient than optical flow or BRIEF+TM. Thus, the most suitable methods for a GCP-enabled case are the interpolation and SRTM registration phases. With accurate RFMs, the SRTM bias removal is expected to be much more effective.

The generated DSM can be improved with other sources of data such as road maps and information on the region (urban, forest, etc.). This is a remote sensing effort that uses the GCP-free DSM, rather than future work on DSM generation, but such work has the potential to improve the DSM accuracy.

REFERENCES

- Abshire, J.B., Sun, X., Riris, H., Sirota, J.M., McGarry, J.F., Palm, S., Yi, D., Liiva, P., 2005, "Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance", *Geophysical Research Letters*, vol. 32, L21S02, doi:10.1029/2005GL024028
- Agrawal, M., Konolige, K., Blas, M. R., 2008, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching", *European Conference on Computer Vision (ECCV 2008)*, part IV, pp. 102–115
- Altamimi, Z., Sillard, P., Boucher, P., 2002, "ITRF2000: A New Release of the International Terrestrial Reference Frame for Earth Science Applications", *Journal of Geophysical Research*, vol. 107, part 10, sect. 2
- Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B., Boucher, C., 2005, "ITRF2005: A New Release of the International Terrestrial Reference Frame Based on Time Series of Station Positions and Earth Orientation Parameters", *Journal of Geophysical Research*, vol. 112, issue B9
- Altamimi, Z., Collilieux, X., Métivier, L., 2011, "ITRF2008: An Improved Solution of the International Terrestrial Reference Frame", *Journal of Geodesy*, vol. 85, no. 8, pp.457-473
- Arias, E. F., 1994, "The Celestial Reference System and Frame of the International Earth Rotation Service", *VLBI TECHNOLOGY: Progress and Future Observational Possibilities*, pp. 313-315
- Aster GDEM Readme File – ASTER GDEM version-1, <http://www.viewfinderpanoramas.org/GDEM/GDEM-README.pdf>, last visit date: September 3, 2012
- Baudoin, A., Schroeder, M., Valorge, C., Bernard, M., Rudowski, V., 2001, "The HRS-SAP initiative: A scientific assessment of the High Resolution Stereoscopic instrument on board of SPOT 5 by ISPRS investigators", *Proc. 10th International Society for Photogrammetry and Remote Sensing*

- (ISPRS) Congress, 12-23 July 2004, Istanbul, Turkey, vol. 25, part B1, pp. 372-378
- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008, "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346--359
- Borkowski, K.M., 1987, "Transformation of Geocentric to Geodetic Coordinates Without Approximations", *Astrophysics and Space Science*, 139 (1987), pp.1-4 <http://www.astro.uni.torun.pl/~kb/Papers/ASS/Geod-ASS.htm> , last visit date: June 9, 2008
- Boucher, C., Altamimi, Z., Sillard, P., 1999, "The International Terrestrial Reference Frame (ITRF97)", *IERS Technical Note 27*, May 1999
- Calonder, M., Lepetit V., Strecha C., Fua P., 2010, "BRIEF- Binary Robust Independent Elementary Features", *ECCV*, September 2010
- Canon Web Site, 2008, "EOS Digital for Professionals Brochure", p. 19, http://www.usa.canon.com/templatedata/pressrelease/20050822_eos5d.html, last visit date: June 4, 2008
- Canny, J., 1986, "A Computational Approach to Edge Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679--714
- Chang, H.C., Li, X., Ge, L., 2010, "Assessment of SRTM, ACE2 and ASTER-GDEM using RTK-GPS". *15th Australasian Remote Sensing & Photogrammetry Conf.*, Alice Springs, Australia, 13-17 September, pp. 271--279 <http://www.15.arspc.com/proceedings>, Last visit date: September 3, 2012
- Clarke, A.C., 1945, "Extra-Terrestrial Relays - Can Rocket Stations Give World-wide Radio Coverage?", *Wireless World*, October 1945, pp. 305-308. http://lakdiva.org/clarke/1945ww/1945ww_oct_305-308.html last visit date: June 9, 2008
- Cook, M.K., Peterson, B.A., Dial, G., Gerlach, F., Hutchins, K., Kudola, R., Bowen, H.S., 2001, "IKONOS Technical Performance Assessment",
- Çiğla, C., Alatan, A.A., 2010, "Efficient Graph-based image segmentation via Speeded-up Turbo Pixels", *IEEE Proceedings of IEEE 17th International Conference on Image Processing*, pp. 3013—3016, September 26-29, 2010, Hong Kong

- Çiğla, C., 2012, "Real-time Stereo to Multi-view Video Conversion, *PhD Thesis*, Middle East Technical University, July 2012
- Dalal, N., BTriggs, B., 2005, "Histograms of Oriented Gradients for Human Detection, *Computer Vision and Pattern Recognition*, San Diego, CA, June 20–25
- d'Angelo, P., Lehner, M., Krauss, T., Hoja, D., Reinartz, P., 2008, "Towards Automated DEM Generation from High Resolution Stereo Satellite Images", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. vol. XXXVII, part B4, Beijing 2008, pp. 1137–1142
- Daugman, J., 1988, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36. no. 7., July 1988, pp. 1169–1179 <http://www.cs.gmu.edu/~zduric/cs774/Papers/Daugman-GaborTransforms.pdf>, last visit date: September 1, 2012
- de Berg, M., Otfried C., van Kreveld M., Overmars M., 2008. "Computational Geometry: Algorithms and Applications". Springer-Verlag. ISBN 978-3-540-77973-5, <http://www.cs.uu.nl/geobook/interpolation.pdf>, last visit date: July 27, 2012
- Delaunay, B., 1934, "Sur la sphère vide", *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7, pp. 793–800
- Di, K., Ma, R., Li, R., 2001, "Deriving 3-D Shorelines from High Resolution IKONOS Satellite Images with Rational Polynomial Functions", *ASPRS Annual Conference*, St. Louis, MO, April 25-27, 2001 (CD-ROM)
- Dial, G., Grodecki, J., 2002a, "Block Adjustment with Rational Polynomial Camera Models", *ACSM-ASPRS 2002 Annual Conference Proceedings*, Washington, DC, April 22-26
- Dial, G., Grodecki, J., 2002b, "IKONOS Accuracy without Ground Control", *Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference Proceedings*
- Dial, G., Bowen, H., Gerlach, F., Grodecki, J., Oleszczuk, R., 2003, "IKONOS satellite, imagery, and products", *Remote Sensing of Environment*, vol. 88 pp.23–36

- DigitalGlobe web page, 2008, "DigitalGlobe: Quickbird",
<http://www.digitalglobe.com/index.php/85/QuickBird>, last visit date:
 June 11, 2008
- Dowman, I. J., Michalis, P., 2003, "Generic Rigorous Model For Along Track
 Stereo Satellite Sensors", *ISPRS Workshop High Resolution Mapping from
 Space 2003*, Hanover, September 19-21
- Duda, R.O., Hart, P.E., 1972, "Use of the Hough Transformation to Detect
 Lines and Curves in Pictures," *Comm. ACM*, vol. 15, pp. 11–15, January
 1972
- EU, EDA, CSG, ESA Joint Task Force, 2010, "Civil-Military Synergies in the
 Field of Earth Observation", Final Report, November 2010,
<http://www.ipi.uni-hannover.de/fileadmin/institut/pdf/dowman.pdf> , last
 visit date: June 11, 2008
- Farr, T. G., *et al.* , 2007, "The Shuttle Radar Topography Mission", *Rev.
 Geophys.*, 45, RG2004, doi:10.1029/2005RG000183
- Försner, W., 1986, "A Feature Based Correspondence Algorithm for Image
 Matching", *ISP Comm. III, Rovaniemi Int. Arch. of Photogrammetry*, vol. 26,
 no. 3, pp. 150-166
- Föstner, W., 1987, "A Fast Operator for Detection and Precise Location of
 Distinct Points, Corners and Centres of Circular Features", *ISPRS
 Intercommission Workshop*, Interlaken, June 1987
- Fraser, C.S. , Hanley, H.B., Yamakawa, T., 2001, "Sub-Metre Geopositioning
 with Ikonos Geo Imagery", *ISPRS Joint Workshop "High Resolution
 Mapping from Space 2001"*, Hanover, Germany, 19-21 September 2001
- Fraser, C.S, 2002, "Prospects for Mapping High-Resolution Satellite
 Imagery", *Proceedings of ACRS 2002*, [http://www.a-a-r-
 s.org/acrs/proceeding/ACRS2002/Papers/VHR02-7.pdf](http://www.a-a-r-s.org/acrs/proceeding/ACRS2002/Papers/VHR02-7.pdf) , last visit date:
 July 24, 2012
- Fraser, C.S., Dare, P.M., Yamakawa, T., 2004, "Digital Surface Modeling from
 SPOT 5 HRS Imagery Using the Affine Projective Model", *International
 Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*,
 vol. 35, part B1, pp. 385–388

- Frey, H., Paul, F., 2011, "On the Suitability of the SRTM DEM and ASTER GDEM for the Compilation of Topographic Parameters in Glacier Inventories", *Int. J. Appl. Earth Observ. Geoinf.*, doi:10.1016/j.jag.2011.09.020
- Gauglitz, S., Höllerer, T., Turk, M., 2011, "Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking", *International Journal of Computer Vision*, vol. 94, pp. 335–360, DOI 10.1007/s11263-011-0431-5
- Gleyzes, J.-P., Meygret, A., Fratter, C., Panem, C., Baillarin, S., Valorge, C., 2003, "SPOT5: system overview and image ground segment", *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS '03*, 2003 vol. 1, 21-25 July 2003 pp. 300 – 302
- GeoEye Web Page, 2007, "IKONOS Relative Spectral Response and Radiometric Calibration Coefficients", <http://www.geoeye.com/products/imagery/ikonos/spectral.htm>, last visit date: June 10, 2008
- German Aerospace Center, 2009, "TerraSAR-X Brochure", http://www.dlr.de/dlr/en/Portaldata/1/Resources/documents/TSX_brosch.pdf, last visit date: June 22, 2012
- Ghadyani, Z., Naseri, S. A., Khiabani, S. A., 2008, "Statistical Study of Space Remote Sensors", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. vol. XXXVII, part B1. Beijing 2008, pp. 567--572
- Gonçalves, J.A., 2008, "Orientation and DEM extraction from ALOS-PRISM Images Using the SRTM-DEM as Ground Control", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, part B1, Beijing 2008, pp. 1177--1182 http://www.isprs.org/proceedings/XXXVII/congress/1_pdf/201.pdf, last visit date: June 20, 2012
- Gonçalves, J.A., 2009, An Empirical Model for Orientation of ALOS-PRISM Images of Level 1B2", *ISPRS Hannover Workshop 2009, High-Resolution Earth Imaging for Geospatial Information* http://www.isprs.org/proceedings/XXXVIII-1-4-7_W5/paper/Goncalves-201.pdf, last visit date: June 20, 2012
- Gonzalez, R. C., Woods, R. E., 2002, "Digital Image Processing", Prentice-Hal

- Grodecki, J., 2001, "IKONOS stereo feature extraction—RPC approach", *Proceedings ASPRS Annual Conference*, St. Louis, 23–27 April, 7 pages (on CD-ROM). http://www.geoeye.com/CorpSite/assets/docs/technical-papers/2001/C_GrodeckiJacek_2001.pdf , last visit date: June 27, 2012
- Grodecki, J., Dial, G., 2003, "Block Adjustment of High-Resolution Satellite Images Described by Rational Polynomials", *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 1, January 2003, pp. 59 – 68
- Gupta, R., Hartley, R.I., 1995, "Camera Estimation for Orbiting Pushbrooms", *Proc. Second Asian Conf. Computer Vision*, Singapore, December 1995
- Gupta, R., Hartley, R.I., 1997, "Linear Pushbroom Cameras", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, September 1997
- Gruen, A.W., 1985, "Adaptive Least Squares Correlation: A Powerful Image Matching Technique", *South African Journal of Photogrammetry, Remote Sensing and Cartography*, vol. 14, no. 3, pp. 175--187
- Gruen, A., Zhang L., Eisenbeiss, H., 2005, "3D Precision Processing of High-Resolution Satellite Imagery", *ASPRS 2005 Annual Conference*, Baltimore, Maryland, March 7-11, 2005
- Hager, J.W., Behensky, J.F., Drew, B.W., 1989, "The Universal Grids: Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS)", *Defence Mapping Agency Technical Manual 8358.2*, 18 September 1989, http://earth-info.nga.mil/GandG/publications/tm8358.2/TM8358_2.pdf , last visit date: June 24, 2012
- Harris, C., Stephens, M., 1988, "A Combined Corner and Edge Detector", *Alvey Vision Conference*, pp. 147--151
- Hartley, R. I., 1993, "Euclidean reconstruction from uncalibrated views", *Proc. of the Second Europe-US Workshop on Invariance*, Ponta Delgada, Azores, pp. 187--202, October 1993
- Hartley, R.I., 1995, "In Defence of the 8-Point Algorithm", *Proc. 5th International Conference on Computer Vision*, Los Alamitos, CA, USA, pp. 1064--1070

- Hartley, R.I., Saxena, T., 1997a "The Cubic Rational Polynomial Camera Model", *Proc. DARPA Image Understanding Workshop 1997*, New Orleans, LA, May 1997, pp. 649--654
- Hartley, R.I., Sturm, P., 1997b, "Triangulation" *Computer Vision and Image Understanding*, vol. 68, pp. 146--157
- Helz, R.L., 2005, "Monitoring Ground Deformation from Space", *US Geological Survey InSAR Factsheet 2005-3025*, July 2005, http://volcanoes.usgs.gov/activity/methods/insar//public_files/InSAR_Fact_Sheet/2005-3025.pdf , last visit date: June 24, 2012
- Horn, B.K.P., Schunck, B.G., 1981, "Determining Optical Flow", *Artificial Intelligence*, vol. 17, pp. 185--203
- Hosseini, M., 2008, "Analysis of Rational Function Dependency to the Height Distribution of Ground Control Points in Geometric Correction of Aerial and Satellite Images", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, part B1, Beijing 2008, pp.1149-1156
- Hu ,Y., Tao, V., Croitoru, A., 2004, "Understanding the Rational Function Model: Methods and Applications, *Proceedings of the XXth International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 35, part B5, 12--23 July 2004, Istanbul, Turkey, unpaginated CD-ROM.
- J Space Systems, 2011, "ASTER Global Digital Elevation Model" <http://www.jspacesystems.or.jp/ersdac/GDEM/E/4.html>, last visit date: September 1, 2012
- Jacobsen, K., 1997, "Geometric Calibration of Space Remote Sensing Cameras for Efficient Processing", *ISPRS International Archives Photogrammetry and Remote Sensing*, vol. 17, no. 1, pp.33--43
- Jacobsen, K., 2002, "Comparison of High Resolution Mapping from Space", *INCA Workshop 2002*, Ahmedabad, India
- Jacobsen, K., 2004, "DEM Generation from Satellite Data", *Remote Sensing in Transition*, pp. 513--525
- Jacobsen, K., Passini, R., 2003, "Accuracy of Digital Orthophotos from High Resolution Space Imagery", *Proceedings of ISPRS Workshop - High*

Resolution Mapping from Space 2003, 04–06 October, Hannover, Germany, unpaginated CD-ROM.

- Jacobsen, K., 2008, "Satellite Image Orientation", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. vol. XXXVII. part B1. Beijing 2008, pp. 703–710
- Jazayeri, I., Fraser, C.S., 2008, "Interest Operators in Close-Range Object Reconstruction", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII. part B5. Beijing 2008, pp 69–74
- Keeratikasikorn, C. , Trisirisatayawong, I., 2008, "Reconstruction of 30m DEM from 90m SRTM SEM with Bicubic Polynomial Interpolation Method", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, part B1, pp.791-794, Beijing 2008
- Kelso, T. S., 1996, "Orbital Coordinate Systems, Part III", *Satellite Times*, vol. 2, no. 3, January/February 1996, pp. 78-79,
<http://www.celestrak.com/columns/v02n03/>, last visit date: June 10, 2012
- Kelso, T. S., 1998, "Basics of the Geostationary Orbit", *Satellite Times*, 4, No 7, pp. 76-77, <http://www.celestrak.com/columns/v04n07/> ,last visit date: June 9, 2012
- Kim, T., 2000, "A Study on the Epipolarity of Linear Pushbroom Images", *Photogrammetric Engineering and Remote Sensing*, vol. 66 no.8, August 2000, pp. 961–966
- Kim, T., Im, Y.-J., 2001, "Automated DEM extraction from the KOMPSAT-1 EOC Images", *ACRS XXII*, November 5-9, 2001, Singapore
- Lee, H.-Y., Kim, T., Park, W., Lee, H.K. 2003, "Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry", *Image and Vision Computing*, Volume 21, Issue 9, 1 September 2003, pp 789–796
- Leprince, S., Barbot, S., Ayoub, F., Avouac, J.P., 2007, "Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to Ground Deformation Measurements", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, June 2007, pp. 1529–1558

- Leutenegger, S., Chli, M., Siegwart, R., 2011, "BRISK: Binary Robust Invariant Scalable Keypoints", *International Conference on Computer Vision (ICCV'11)*
- Lowe, D. G., 2004, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110
- Lucas, B. D., Kanade, T., 1981, "An Iterative Image Registration Technique with an Application to Stereo Vision", *Proceedings of Imaging Understanding Workshop*, pp. 121--130
- Mair, E., Hager, G. D., Burschka, D., Suppa, M., Hirzinger, G., 2010, "Adaptive and generic corner detection based on the accelerated segment test". *Proceedings of the European Conference on Computer Vision (ECCV 2010)*
- Matas, J., Chum, O., Urban, M., Pajdla, T., 2002, "Robust widebaseline stereo from maximally stable extremal regions", *BMVC 2002*, September 2002, pp. 384--393
- Michalis, P., Dowman, I., 2004, "A Rigorous Model and DEM Generation for SPOT5 –HRS", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 35, no.B1, pp. 410--415
- Mohr, P. J., Taylor, B. N., Newell, D. B., 2006, *CODATA Recommended Values of the Fundamental Physical Constants: 2006*, p.59
<http://physics.nist.gov/cuu/Constants/codata.pdf>, last visit date: 6th June, 2008
- Morgan, M., 2004, "Epipolar Resampling of Linear Array Scanner Scenes", *UCGNE Reports, Number 20193*, May 2004, 187 pages
- Moritz, H., 1992, "Geodetic Reference System 1980", *The Geodesists Handbook 1992 (Bulletin Geodesique)*, vol. 66, no. 2, 1992,
http://www.gfy.ku.dk/~iag/HB2000/part4/grs80_corr.htm , last visit date: June 11, 2008
- NIMA (National Imagery and Mapping Agency), 2000, "Department of Defence World Geodetic System 1984, Its Definition and Relationships with Local Geodetic Systems", *Technical Report NIMA TR8350.2*, 3rd Ed. 3 January 2000,
<http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf> , last visit date: June 22, 2012

- Ojala, T., Pietikäinen, M., Harwood D., 1994, "Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions", *Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994)*, vol. 1, pp. 582 - 585.
- Osawa, Y. 2004, "Optical And Microwave Sensors On Japanese Mapping Satellite – ALOS", *Proc. 10th International Society for Photogrammetry and Remote Sensing (ISPRS) Congress*, 12-23 July, 2004, Istanbul, Turkey, pp. 309-312
- Otto, G.P., Chau, T.K.W., 1988. "A Region-Growing Algorithm for Matching of Terrain Images". *Proc. 4th Alvey Vision Club*, University of Manchester, UK, 31 Aug. 2 Sept 1988.
- Perry, R., 1973, "A History of Satellite Reconnaissance Volume I -Corona", *Central Intelligence Agency*, October 1973, 262 pages, http://www.foia.cia.gov/docs/DOC_0000940279/DOC_0000940279.pdf, last visit date: July 9, 2012
- Piwowar, J.M., 2009, "IKONOS Satellites", <http://uregina.ca/piwowarj/Satellites/IKONOS.html>, last visit date: June 9, 2012
- Poli, D., 2003, "Georeferencing of Moms-02 And MISR Stereoimages With Strict Sensor Model", *ISPRS Workshop High Resolution Mapping from Space 2003*, Hannover, 4-6 October 2003
- Poli, D., 2004a, "Orientation of Satellite and Airborne Imagery From Multi-Line Pushbroom Sensors With A Rigorous Sensor Model", *Proc. 10th International Society for Photogrammetry and Remote Sensing (ISPRS) Congress*, 12-23 July, 2004, Istanbul, Turkey, pp.130-135,
- Poli, D., Li, Z., Gruen, A., 2004b, "SPOT-5/HRS Stereo Images Orientation and Automated DSM Generation", *Proc. 10th International Society for Photogrammetry and Remote Sensing (ISPRS) Congress*, 12-23 July, 2004, Istanbul, Turkey, pp.421-432,
- Poli, D, 2007, "Modelling of spaceborne linear array sensors", *PhD Thesis, published by Institute of Geodesy and Photogrammetry Swiss Federal Institute of Technology (ETH), Zurich*, 199 pages, ISBN 3-906467-55-4

- Poli, D., Toutin, T., 2012, "Review of Developments in Geometric Modelling for High Resolution Satellite Pushbroom Sensors", *The Photogrammetric Record*, vol. 27, no. 137, March 2012, pp. 58–73
- Prewitt, J.M.S., 1970, "Object Enhancement and Extraction", *Picture Processing and Psychopictorics*, 1970
- Riazanoff, S., 2004, "SPOT 123-4-5 Geometry Handbook", Gael Consultant, GAEL-P135-DOC-001, 82 pages, <http://www-igm.univ-mlv.fr/~riazano/publications/GAEL-P135-DOC-001-01-04.pdf> , last visit date: July 25, 2012
- Richards, J. A., Jia, X., 2006, "Error Correction and Registration of Image Data", *Remote Sensing Digital Image Analysis-An Introduction*, chapter 2, Springer-Verlag, pp. 46--54, ISBN-10 3-540-25128-6
- Roberts, L.G., 1963, "Machine Perception of Three-Dimensional Solids", *MIT Lincoln Laboratory Technical Report #315*, 22 May 1963. <http://www.packet.cc/files/mach-per-3D-solids.html>, last visit date: September 1, 2012
- Rodriguez, E., Morris, C.S., Belz, J.E., Chapin, E.C., Martin, J.M., Daffer, W., Hensley, S., 2005, "An Assessment of the SRTM Topographic Products", *Technical Report JPL D-31639*, Jet Propulsion Laboratory, Pasadena, California, 143 pp., http://www2.jpl.nasa.gov/srtm/SRTM_paper.pdf, last visit date: June 26, 2012
- Rosten, E., Drummond, T., 2006, "Machine learning for high-speed corner detection", *ECCV 2006*, May 2006, pp. 430–443
- Rosten, E., Drummond, T., 2005, "Fusing Points and Lines for High Performance Tracking", *ICCV 2005*, October 2005, pp. 1508–1511
- Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011, "ORB: an efficient alternative to SIFT or SURF", *ICCV 2011*, 6-13 November 2011, pp. 2564--2571
- Satellite Imaging Corp. Web Site, "Satellite Imaging Sensors", <http://www.satimagingcorp.com/satellite-sensors.html> , last visit date: June 24, 2012

- Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., DiMarzio, J. P. , 2005, "Overview of the ICESat Mission", *Geophys. Res. Lett.*, 32, L21S01, doi:10.1029/2005GL024009.
<http://icesat.gsfc.nasa.gov/icesat/publications/GRL/schutz-1.pdf>, last visit date: June 27, 2012
- Shewchuk J. R., 1996, "Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator", *Applied Computational Geometry: Towards Geometric Engineering, Lecture Notes in Computer Science*, vol. 1148, pp. 203-222, Springer-Verlag, Berlin, May 1996.
- Shi, J., Tomasi, C., 1994, "Good Features to Track", *CVPR94*, Seattle, June 1994, <http://www.ai.mit.edu/courses/6.891/handouts/shi94good.pdf> , last visit date: August 28, 2012
- Silveira, M. T., Feitosa, R. Q., Jacobsen K., Brito, J. L. N. S., Heckel, Y., 2008, "A Hybrid Method for Stereo Image Matching", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII. part B1. Beijing 2008, pp. 895—900
- Simpson, D.G., "Optics of the Hubble Space Telescope", May 2010. <http://www.pgccphy.net/ref/hst-optics.pdf> , last visit date: July 9, 2012
- Smith, S. M., Brady, J. M., 1997, "SUSAN - A New Approach to Low Level Image Processing", *International Journal of Computer Vision*, vol. 23, no 1, pp. 45—78
- Space Telescope Science Institute (STScI), 2011, "Hubble Space Telescope Primer for Cycle 20", December 2011, <http://www.stsci.edu/hst/proposing/documents/cp/primer.pdf> , last visit date: July 9, 2012
- Tachikawa, T., Hato, M., Kaku, M., Iwasaki, A., 2011, "The Characteristics of ASTER GDEM Version 2", *IGARSS*, July 2011.
- Tao, C.V., Hu, Y., 2002, "3D Reconstruction Models Based on the Rational Function Model", *Photogrammetric Engineering & Remote Sensing*, Vol. 68, No. 7, July 2002, pp. 705-714.
- The Defence Mapping Agency, 1983, "Geodesy for the Layman", *DMA Technical Report DMA TR 80-003*, December 1983, 103 pages
http://www.ngs.noaa.gov/PUBS_LIB/Geodesy4Layman/geo4lay.pdf, last visit date: August 30, 2012

- Tola, E., Lepetit, V., Fua, P., 2010, "DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, May 2010
- Toutin, T., Gray, L., 2000, "State-of-the-art of Elevation Extraction from Satellite SAR Data", *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 55, pp. 13--33
- Toutin, T., 2001, "Elevation Modelling from Satellite Visible and Infrared (VIR) Data", *International Journal of Remote Sensing*, 2001, vol. 22, no. 6, pp. 1097--1125
- Toutin, T., 2004, "Review article: Geometric processing of remote sensing images: models, algorithms and methods", *International Journal of Remote Sensing*, vol. 25, no. 10, pp. 1893-1924
- Toutin, T., 2009, "Fine Spatial Resolution Optical Sensors", *Chapter 8 in SAGE Handbook of Remote Sensing* (Eds.T. A. Warner, M. D. Nellis & G. M. Foody). SAGE, London, UK., pp. 108--122.
- Toutin, T., 2012, "Radarsat-2 DSM Generation With New Hybrid, Deterministic, and Empirical Geometric Modeling Without GCP", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, No. 5, May 2012, pp. 2049--2055
- Triangle web page, "Triangle: A Two-Dimensional Quality Mesh Generator and Delaunay Triangulator",
<http://www.cs.cmu.edu/~quake/triangle.html>, last visit date: August 20, 2012
- Tsutsui, K., Rokugawa, S., Nakagawa, H., Miyazaki, S., Cheng, C.-T. , Shiraishi, T., Yang, S.-D., 2007, "Detection and Volume Estimation of Large-Scale Landslides Based on Elevation-Change Analysis Using DEMs Extracted From High-Resolution Satellite Stereo Imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, June 2007, pp. 1681--1696
- Tuytelaars, T., and Mikolajczyk, K., 2007, "Local Invariant Feature Detectors: A Survey", *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3 , 2007, pp.177--280
- US Department of Defence, 1997, "Department of Defense World Geodetic System 1984, Its Definition and Relationships with Local Geodetic

- Systems", *NIMA Technical Report TR8350.2*, Third Edition, 4 July 1997 (updated on June 23, 2004)
<http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf> , last visit date: June 9, 2008
- UCS web page, "UCS Satellite Database",
http://www.ucsusa.org/nuclear_weapons_and_global_security/space_weapons/technical_issues/ucs-satellite-database.html , last visit date: July 9, 2012
- USGS EO-1 website, 2007, "Sensors – Hyperion",
<http://eo1.usgs.gov/hyperion.php> , last visit date: June 10, 2008
- Wang, Y., 1999, "Automated Triangulation of Linear Scanner Imagery", 1999. *Proc. Joint ISPRS Workshop "Sensors and Mapping from Space,"* Hannover, 27– 30 September. Institute of Photogrammetry & Geoinformation, University of Hannover, 5 pages (on CD)
www.ipi.uni-hannover.de/fileadmin/institut/pdf/wang.pdf, last visit date: June 6, 2012
- Wang, Y., Hu, F., Li, J., 2010, "Epipolar Arrangement of Satellite Imagery by Projection Trajectory Simplification", *The Photogrammetric Record*, vol. 25, no. 132, pp. 422--436, doi: 10.1111/j.1477-9730.2010.00602.x
- Yamakawa, T., Fraser, C.S., 2004, "The Affine Projection Model For Sensor Orientation: Experiences With High-Resolution Satellite Imagery", *Proc. 10th International Society for Photogrammetry and Remote Sensing (ISPRS) Congress*, 12-23 July, 2004, Istanbul, Turkey, pp.142--147, <http://www.isprs.org/istanbul2004/comm1/papers/27.pdf> last visit date: June 12, 2008
- Yoder, C. F., 1995, "Astrometric and Geodetic Properties of Earth and the Solar System", in T. J. Ahrens: *Global Earth Physics: A Handbook of Physical Constants*. Washington: American Geophysical Union. ISBN 0875908519, pp. 1–31
- Zhang, L, 2005, "Automatic Digital Surface Model (DSM) Generation from Linear Array Images", *PhD Thesis*, published by *Institute of Geodesy and Photogrammetry Swiss Federal Institute of Technology (ETH), Zurich*, 199 pages, ISBN 3-906467-55-4

Zhao D., Yuan, X., Liu, X., 2008, "Epipolar Line Generation from IKONOS Imagery Based on Rational Function Model", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. vol. XXXVII. part B4. Beijing 2008, pp. 1293–1298

APPENDIX A

EMPIRICAL MODELS FROM SRTM REGISTRATION

The empirical models require ground control data for determining the model parameters. Obviously, the quality of the ground control data has a significant effect on the model accuracy. The RPCs provided with the satellite images are generated from the undisclosed physical models by the image providers. However, studies on simpler models such as APM and DLT are still being published and it is shown that these simpler models are also appropriate for DSM generation, provided that sufficiently many GCPs are available (*Yamakawa 2004, Jacobsen 2008*). Simpler empirical models are not tested when less accurate ground control data are available.

With the SRTM registration proposed in this study, all points in the image are assigned geodetic coordinates with some moderate accuracy. The accuracy is improved with bias reduction using accurate correspondences. After SRTM registration, moderate ground control data are available for all pixels in the image instead of a few sparsely distributed but accurate GCPs.

The major problem for simple models is that, the model assumptions are invalidated by the orbital motion of the satellite. With the availability of many ground control data; it is possible to divide the satellite image into row chunks and fit a separate model for each row chunk to obtain a piecewise

projection model. Here, we investigate the achievable accuracy for such models.

Two projection models are tested: The piecewise pinhole model (P-PM) and the piecewise DLT (P-DLT) model. For each model, different chunk sizes are used and the generated models are tested at the available GCPs. The following experiments are conducted:

- i) SRTM registration bias is manually corrected in the object domain and used as the ground control data to obtain the piecewise model parameters. This experiment is performed to observe the best achievable accuracy for the piecewise models when SRTM registration is used.
- ii) Biased SRTM registration is used as the ground control data to obtain the piecewise model parameters.
- iii) The result of the bias-correction step (Section 4.2.1) is used as the ground control data to obtain the piecewise model parameters. This is the best achievable accuracy for SRTM-based piecewise models.

The test is performed on the IKONOS image for the Hobart area, where 122 GCPs are available. The errors are computed only in the row chunks that contain GCPs. The results are presented in the tables below.

Table A.1. Image domain projection errors for the piecewise models when the SRTM registration bias is manually removed in the object domain. All values are rms errors in pixels, L is the number of rows in a chunk

L	5	50	100	1000	2000	4000	6000
P-PM	2.3	2.17	2.23	2.1	2.16	2.13	2.45
P-DLT	2.17	2.19	2.16	2.08	2.1	2.06	2.23

Table A.2. Image domain projection errors for the piecewise models without bias removal for the SRTM registration. All values are rms errors in pixels, L is the number of rows in a chunk

L	5	50	100	1000	2000	4000	6000
P-PM	7.11	7.1	7.13	7.07	7.14	6.97	7.29
P-DLT	7.06	7.11	7.11	7.05	7.1	7.06	7.16

Table A.3. Image domain projection errors for the piecewise models when the SRTM registration bias is removed automatically in the object domain. All values are rms errors in pixels, L is the number of rows in a chunk

L	5	50	100	1000	2000	4000	6000
P-PM	3.01	2.92	2.96	2.88	2.94	2.87	3.19
P-DLT	6.22	2.92	2.96	2.88	2.93	2.87	3.1

As seen in the results, the SRTM bias reduction provides accuracy figures that are very close to the the best obtainable accuracy for piecewise models obtained by SRTM registration.

The results also indicate that, the piecewise models can be used with different values of L, but using less than 50 or more than 4000 rows is not advised for narrow field-of view satellites, such as IKONOS.

Of course, the image domain projection errors are larger than those of a GCP-based model. Still, the results show that, piecewise empirical models provide moderate accuracy in the absence of GCPs. These models can be used when the GCPs are unavailable and the accuracy needs are moderate.

Another use of this approach is the following: The projection errors presented above have strong bias terms (the error is mostly bias), which can be corrected with a single GCP. Thus, although the DLT or the pinhole model

parameters cannot be computed with a single GCP, when the SRTM registration is available, it is possible to obtain good projection models even with a single GCP.

Note that, the SRTM registration is obtained from third order RFMs, and if the RFMs are present, the need for simpler models is questionable. However, the results of piecewise empirical models are better than the original biased RFMs, thanks to the object domain bias correction for the SRTM registration. The bias-corrected SRTM cannot be used to improve the original third order RFMs, because the object domain bias correction terms undergo third order polynomials in the original RFMs and distort the cross-terms in the polynomials.

The piecewise empirical model is a side product of the thesis study and need further investigation.

APPENDIX B

JACOBIAN IN THE RPC RECONSTRUCTION

Here, we provide the complete and detailed formulas for every entry of the Jacobian matrix that is used in RPC reconstruction with Levenberg-Marquardt algorithm. The Jacobian matrix is:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial u_1}{\partial \phi} & \frac{\partial u_1}{\partial \lambda} & \frac{\partial u_1}{\partial h} \\ \frac{\partial v_1}{\partial \phi} & \frac{\partial v_1}{\partial \lambda} & \frac{\partial v_1}{\partial h} \\ \frac{\partial u_2}{\partial \phi} & \frac{\partial u_2}{\partial \lambda} & \frac{\partial u_2}{\partial h} \\ \frac{\partial v_2}{\partial \phi} & \frac{\partial v_2}{\partial \lambda} & \frac{\partial v_2}{\partial h} \end{bmatrix} \quad (\text{B.1})$$

As noted earlier, the normalization (offset and scale) parameters are provided together with the RPCs, by the image vendors. The object domain normalization maps the geodetic latitude, longitude and height to the [-1, +1] interval.

The normalization equations for image 1 are:

$$\begin{aligned}
X_1 &= \frac{\phi - \phi_{1,offset}}{\phi_{1,scale}} \\
Y_1 &= \frac{\lambda - \lambda_{1,offset}}{\lambda_{1,scale}} \\
Z_1 &= \frac{h - h_{1,offset}}{h_{1,scale}}
\end{aligned} \tag{B.2}$$

The object domain normalization equations for image 2 are:

$$\begin{aligned}
X_2 &= \frac{\phi - \phi_{2,offset}}{\phi_{2,scale}} \\
Y_2 &= \frac{\lambda - \lambda_{2,offset}}{\lambda_{2,scale}} \\
Z_2 &= \frac{h - h_{2,offset}}{h_{2,scale}}
\end{aligned} \tag{B.3}$$

The image domain de-normalization equations for image 1 and image 2 are:

$$\begin{aligned}
u_1 &= y_1 * u_{1,scale} + u_{1,offset} \\
v_1 &= x_1 * v_{1,scale} + v_{1,offset} \\
u_2 &= y_2 * u_{2,scale} + u_{2,offset} \\
v_2 &= x_2 * v_{2,scale} + v_{2,offset}
\end{aligned} \tag{B.4}$$

The projection functions are defined with the normalized coordinates and are as follows:

$$\begin{aligned}
y_1 &= \frac{P_{11}(X_1, Y_1, Z_1)}{P_{12}(X_1, Y_1, Z_1)} & , & & x_1 &= \frac{P_{13}(X_1, Y_1, Z_1)}{P_{14}(X_1, Y_1, Z_1)} \\
y_2 &= \frac{P_{21}(X_2, Y_2, Z_2)}{P_{22}(X_2, Y_2, Z_2)} & , & & x_2 &= \frac{P_{23}(X_2, Y_2, Z_2)}{P_{24}(X_2, Y_2, Z_2)}
\end{aligned} \tag{B.5}$$

Numerator for the rows of image 1:

$$\begin{aligned}
P_{11}(X, Y, Z) &= a_0 + a_1X_1 + a_2Y_1 + a_3Z_1 + a_4X_1^2 \\
&+ a_5X_1Y_1 + a_6X_1Z_1 + a_7Y_1^2 + a_8Y_1Z_1 \\
&+ a_9Z_1^2 + a_{10}X_1^3 + a_{11}X_1^2Y_1 + a_{12}X_1^2Z_1 \\
&+ a_{13}Y_1^2X_1 + a_{14}Y_1^2Z_1 + a_{15}Z_1^2X_1 \\
&+ a_{16}Y_1^3 + a_{17}Y_1Z_1^2 + a_{19}Z_1^3
\end{aligned} \tag{B.6}$$

We don't feel the necessity to write down all 8 polynomials separately, since they are all in the same form. The only difference is the values of the coefficients and the normalized variables for image 1 and image 2. It is sufficient to assign different letters to the polynomial coefficients: The resultant coefficient labels are given below:

Table B.1. Labelling for the projection polynomial coefficients

Image 1			Image 2		
Polynomial	Coefficients	Variables	Polynomial	Coefficients	Variables
P ₁₁	a ₀ ... a ₁₉	X ₁ , Y ₁ , Z ₁	P ₂₁	e ₀ ... e ₁₉	X ₂ , Y ₂ , Z ₂
P ₁₂	b ₀ ... b ₁₉	X ₁ , Y ₁ , Z ₁	P ₂₂	f ₀ ... f ₁₉	X ₂ , Y ₂ , Z ₂
P ₁₃	c ₀ ... c ₁₉	X ₁ , Y ₁ , Z ₁	P ₂₃	g ₀ ... g ₁₉	X ₂ , Y ₂ , Z ₂
P ₁₄	d ₀ ... d ₁₉	X ₁ , Y ₁ , Z ₁	P ₂₄	h ₀ ... h ₁₉	X ₂ , Y ₂ , Z ₂

As explained in section 5.3.2, the top-left entry of the Jacobian is calculated from the following relation.

$$\frac{\partial u_1}{\partial \phi} = \frac{\partial u_1}{\partial y_1} \left(\frac{\partial y_1}{\partial X_1} \frac{\partial X_1}{\partial \phi} + \frac{\partial y_1}{\partial Y_1} \frac{\partial Y_1}{\partial \phi} + \frac{\partial y_1}{\partial Z_1} \frac{\partial Z_1}{\partial \phi} \right) \quad (\text{B.7})$$

As seen clearly in the object domain normalization equations;

- X is independent of ϕ and h
- Y is independent of λ and h
- Z is independent of ϕ and λ

Consequently, the following partial derivatives are zero:

$$\frac{\partial X_1}{\partial \lambda} = 0, \quad \frac{\partial X_1}{\partial h} = 0, \quad \frac{\partial Y_1}{\partial \phi} = 0, \quad \frac{\partial Y_1}{\partial h} = 0, \quad \frac{\partial Z_1}{\partial \phi} = 0, \quad \frac{\partial Z_1}{\partial \lambda} = 0$$

Thus, in (B.7), the summation in the parenthesis has only one non-zero term. The equation reduces to:

$$\frac{\partial u_1}{\partial \phi} = \frac{\partial u_1}{\partial y_1} \frac{\partial y_1}{\partial X_1} \frac{\partial X_1}{\partial \phi} \quad (\text{B.8})$$

The first and the last terms of equation (B.8) are very easy to compute; from (B.2), (B.3) and (B.4), and they appear as constants:

$$\frac{\partial u_1}{\partial y_1} = u_{1,scale} \quad \frac{\partial u_2}{\partial y_2} = u_{2,scale} \quad (\text{B.9})$$

$$\frac{\partial v_1}{\partial y_1} = v_{1,scale} \quad \frac{\partial v_2}{\partial y_2} = v_{2,scale}$$

$$\frac{\partial X_1}{\partial \phi} = \frac{1}{\phi_{1,scale}} \quad \frac{\partial X_2}{\partial \phi} = \frac{1}{\phi_{2,scale}}$$

$$\frac{\partial Y_1}{\partial \phi} = \frac{1}{\lambda_{1,scale}} \quad \frac{\partial Y_2}{\partial \phi} = \frac{1}{\lambda_{2,scale}} \quad (\text{B.10})$$

$$\frac{\partial Z_1}{\partial h} = \frac{1}{h_{1,scale}} \quad \frac{\partial Z_2}{\partial h} = \frac{1}{h_{2,scale}}$$

$\frac{\partial y_1}{\partial X_1}$ is obtained using the chain rule for the y_1 function in (B.5):

$$\frac{\partial y_1}{\partial X_1} = \frac{P_{12}(X_1, Y_1, Z_1) \left. \frac{\partial P_{11}}{\partial X_1} \right|_{(X_1, Y_1, Z_1)} - P_{11}(X_1, Y_1, Z_1) \left. \frac{\partial P_{12}}{\partial X_1} \right|_{(X_1, Y_1, Z_1)}}{P_{12}^2(X_1, Y_1, Z_1)} \quad (\text{B.11})$$

From (B.6), the partial derivatives of the polynomials are easy to obtain:

$$\begin{aligned} \frac{\partial P_{ij}(X_i, Y_i, Z_i)}{\partial X_i} &= \omega_1 + 2\omega_4 X_i + \omega_5 Y_i + \omega a_6 Z_i \\ &+ 3\omega_{10} X_i^2 + 2\omega_{11} X_i Y_i + 2\omega_{12} X_i Z_i + \omega_{13} Y_i^2 + \omega_{15} Z_i^2 \end{aligned} \quad (\text{B.12})$$

$$\begin{aligned} \frac{\partial P_{ij}(X_i, Y_i, Z_i)}{\partial Y_i} &= \omega_2 Y_i + \omega a_5 X_i + 2\omega_7 Y_i + \omega_8 Z_i \\ &+ \omega_{11} X_i^2 + 2\omega_{13} Y_i X_i + 2\omega_{14} Y_i Z_i + 3\omega_{16} Y_i^2 + \omega_{17} Z_i^2 \end{aligned} \quad (\text{B.13})$$

$$\begin{aligned} \frac{\partial P_{ij}(X_i, Y_i, Z_i)}{\partial Z_i} &= \omega_3 Z_i + \omega_6 X_i + \omega_8 Y_i + 2\omega_9 Z_i \\ &+ \omega_{12} X_i^2 + \omega_{14} Y_i^2 + 2\omega_{15} Z_i X_i + 2\omega_{17} Y_i Z_i + 3\omega_{19} Z_i^2 \end{aligned} \quad (\text{B.14})$$

Here, the subscript i shows the image index, and j is 1, 2, 3 or 4. For each (i, j) pair, the coefficients (ω_{ij}) are taken from Table B.1

The resultant Jacobian matrix entries are obtained by putting equations (B.12), (B.13), and (B.14) into (B.11) correctly and then putting (B.9), (B.10) and (B.11) into (B.8). The short forms of the matrix entries are presented below.

Table B.1. The entries in the Jacobian matrix

$\frac{\partial u_1}{\partial \phi} = \frac{\partial u_1}{\partial y_1} \frac{\partial y_1}{\partial X_1} \frac{\partial X_1}{\partial \phi}$	$\frac{\partial u_1}{\partial \lambda} = \frac{\partial u_1}{\partial y_1} \frac{\partial y_1}{\partial Y_1} \frac{\partial Y_1}{\partial \lambda}$	$\frac{\partial u_1}{\partial h} = \frac{\partial u_1}{\partial y_1} \frac{\partial y_1}{\partial Z_1} \frac{\partial Z_1}{\partial h}$
$\frac{\partial v_1}{\partial \phi} = \frac{\partial v_1}{\partial x_1} \frac{\partial x_1}{\partial X_1} \frac{\partial X_1}{\partial \phi}$	$\frac{\partial v_1}{\partial \lambda} = \frac{\partial v_1}{\partial x_1} \frac{\partial x_1}{\partial Y_1} \frac{\partial Y_1}{\partial \lambda}$	$\frac{\partial v_1}{\partial h} = \frac{\partial v_1}{\partial x_1} \frac{\partial x_1}{\partial Z_1} \frac{\partial Z_1}{\partial h}$
$\frac{\partial u_2}{\partial \phi} = \frac{\partial u_2}{\partial y_2} \frac{\partial y_2}{\partial X_2} \frac{\partial X_2}{\partial \phi}$	$\frac{\partial u_2}{\partial \lambda} = \frac{\partial u_2}{\partial y_2} \frac{\partial y_2}{\partial Y_2} \frac{\partial Y_2}{\partial \lambda}$	$\frac{\partial u_2}{\partial h} = \frac{\partial u_2}{\partial y_2} \frac{\partial y_2}{\partial Z_2} \frac{\partial Z_2}{\partial h}$
$\frac{\partial v_2}{\partial \phi} = \frac{\partial v_2}{\partial x_2} \frac{\partial x_2}{\partial X_2} \frac{\partial X_2}{\partial \phi}$	$\frac{\partial v_2}{\partial \lambda} = \frac{\partial v_2}{\partial x_2} \frac{\partial x_2}{\partial Y_2} \frac{\partial Y_2}{\partial \lambda}$	$\frac{\partial v_2}{\partial h} = \frac{\partial v_2}{\partial x_2} \frac{\partial x_2}{\partial Z_2} \frac{\partial Z_2}{\partial h}$

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Ataseven, Yoldaş
Nationality: Turkish (TC)
Date and Place of Birth: 14 January 1980, Karaman
Marital Status: Single
Phone: +90 312 592 23 67
E-mail: yoldas@eee.metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Electrical and Electronics Eng.	2005
BS	METU Electrical and Electronics Eng.	2002
High School	Ankara Science High School	1998

WORK EXPERIENCE

Year	Place	Enrollment
2008-Present	ASELSAN A.Ş.	Senior Systems Engineer
2002-2008	METU EEE Dept.	Research Assistant
2001 August	ESDAŞ	Intern Engineering Student
2000 July	ASELSAN A.Ş.	Intern Engineering Student

FOREIGN LANGUAGES

Advanced English, Fluent Spanish

PUBLICATIONS

1. Ataseven Y., Alatan A. A., 2010, "SRTM Registration for Electro-optic Satellite Images without GCP", In: Paparoditis N., Pierrot-Deseilligny M., Mallet C., Tournaire O. (Eds), *IAPRS*, Vol. XXXVIII, Part 3A pp. 2004-2008, Saint-Mandé, France, September 1-3, 2010
2. Ataseven Y., Alatan A. A., 2010, "Fast Stereo Pair Generation using SRTM Registration and Optical Flow", *ISPRS İstanbul Workshop 2010 on Modeling Optical Airborne and Spaceborne Sensors*, WG I/4 Oct. 11-13, IAPRS Vol. XXXVIII, part 1/W4, http://www.isprs.org/proceedings/XXXVIII/1-W17/16_Ataseven.pdf
3. Erdener A., Arı E.O., Ataseven Y., Deniz B., İnce K.G., Kazancıoğlu U., Açıkgöz-Kopanoğlu T., Koray T., Koşaner, K.M., Özgür A., Çakıcı-Özkök Ç., Soncul T., Şirin H.O., Yakın İ., Biddlestone S., Fu L., Kurt A., Özgüner Ü., Redmill K., Aytekin Ö., Ulusoy İ., 2010, "Team Cappadocia Design for MAGIC 2010", *Land Warfare Conference 2010*, Brisbane Australia, November 2010
4. Ataseven Y., Akalın-Acar Z., Acar C. E., Gençer, N. G., 2008, "Parallel Implementation of the Accelerated BEM Approach for EMSI of the Human Brain", *Medical and Biological Engineering and Computing*, vol. 46, pp. 671-679

HOBBIES

Photography, Short Film Making, Bağlama, Solar Eclipses