BUILDING DETECTION IN OFF-NADIR VERY HIGH RESOLUTION SATELLITE IMAGES BASED ON STEREO 3D INFORMATION

ALAELDIN SULIMAN

January 2017

BUILDING DETECTION IN OFF-NADIR VERY HIGH RESOLUTION SATELLITE IMAGES BASED ON STEREO 3D INFORMATION

Alaeldin Suliman

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

January 2017

PREFACE

This technical report is a reproduction of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geodesy and Geomatics Engineering, January 2017. The research was supervised by Dr. Yun Zhang, and funding was provided in part by the Libyan Ministry of Higher Education and Scientific Research (LMHEAR), and the Canada Research Chairs (CRC) Program.

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

ABSTRACT

Mapping or updating maps of urban areas is crucial for urban planning and management. Since buildings are the main objects in urban environments, building roof detection is an important task in urban mapping. The ideal geo-spatial data source for mapping building information is very high resolution (VHR) satellite images. On the other hand, because buildings are elevated objects, incorporating their heights in building detection can significantly improve the accuracy of the mapping. The most cost-effective source for extracting the height information is stereo VHR satellite images that can provide two types of stereo 3D information: elevation and disparity. However, most VHR images are acquired off-nadir. This acquisition type causes building leaning in the images and creates major challenges for the incorporation of building height information into roof detection. Thus, this PhD research focuses on finding solutions to mitigate the problems associated with 3D-supported building detection in off-nadir VHR satellite images. It also exploits the potential of extracting disparity information from off-nadir image pairs to support building detection.

In the research, several problems associated with building leaning need to be solved, such as building roof offsetting from its footprint, object occlusion, and building façades. Moreover, the variation of the roofs offsets based on the building heights. While the offsets of building roof create difficulties in the co-registration between image and elevation data, the building façades and occlusions create challenges in automatically finding matching points in off-nadir image pairs. Furthermore, due to the variation in

building-roof offsets, the mapped roofs extracted from off-nadir images cannot be directly geo-referenced to existing maps for effective information integration.

In this PhD dissertation, all of the above identified problems are addressed in a progressively improving manner (i.e., solving the problems one after another while improving the efficiency) within the context of 3D-supported building detection in off-nadir VHR satellite images. Firstly, an image-elevation co-registration technique is developed that is more efficient than the currently available techniques. Secondly, the computation cost is then reduced by generating disparity information instead of the traditional elevation data. This allows bypassing a few time-consuming steps of the traditional method. Thirdly, the disparity generation is then extended from using one pair of off-nadir images to using multiple pairs for achieving an enriched disparity map. Finally, the enriched disparity maps achieved are then used to efficiently derive elevations that are directly co-registered with pixel-level accuracy to the selected reference image. Based on these disparity-based co-registered elevations, building roofs are successfully detected and accurately geo-referenced to existing maps.

The outcome of this PhD research proved the possibility of using off-nadir VHR satellite images for accurate urban building detection. It significantly increases the data source scope for building detection since most (> 95%) of VHR satellite images are off-nadir and traditional methods cannot effectively handle off-nadir images.

DEDICATION

IN MEMORY OF MY MOTHER AND FATHER

TO MY WIFE

WITH LOVE AND ETERNAL APPRECIATION

ACKNOWLEDGEMENTS

Table of Contents

List of Tables

List of Figures

xiii

xv

List of Symbols, Nomenclature or Abbreviations

2D   – Two Dimensional

3D   – Three Dimensional

VHR   – Very High Resolution

MVS   – Multi-View Stereo images

MVIS – Multi-View Inline Stereo images

GIS   – Geographic Information Systems

OBIA – Object-based Image Analysis

DEM   – Digital Elevation Model

DSM   – Digital Surface Model

DTM   – Digital Terrain Model

nDSM – normalized Digital Surface Model

LoS-DSM – Line-of-Sight Digital Surface Model

DSM' – Registered line-of-sight DSM

DTM' – Extracted DTM from the registered line-of-sight DSM.

$nDSM'$– normalized line-of-sight Digital Surface Model

SDM   – Surface Disparity Map

TDM   – Terrain Disparity Map

nSDM – normalized Surface Disparity Map

RFM   – Rational Function Model

RPC   – Rational polynomial Coefficient

GCPs – Ground Control Points

CPs   – Check Points

CNL – Cognition Network Language

I-R – 2D Image Registration

O-R – Orthorectification

P-M – Proposed Method

RMSE – Root Mean Square Error

NDVI – Normalized Difference Vegetation Index

Comp. – Completeness accuracy measure

Corr. – Correctness accuracy measure

OQ. – Overall Quality accuracy measure

TP – True Positive

FP – False Positive

FN – False Negative

RMAD – Registration-based Mapping of Aboveground Disparities

DPP – Double Projection Planes method

DECR – Disparity-based Elevation Co-registration Method

θ – Rotation angle to the epipolar direction

$\Delta X_{ortho}^{k\prime}$ – Shift to the correct orthographic location in the epipolar direction

$X_{ortho}^{k\prime}$ – The correct orthographic location in the epipolar direction

Z_{avg} – The average terrain elevation of the imaged area

Chapter 1: INTRODUCTION

This PhD dissertation extends previous research on urban mapping using very high resolution (VHR) optical satellite imagery. Buildings are the main objects in urban environments, and VHR satellite optical images provide the ideal geo-spatial data for mapping building information. Hence, this dissertation presents research on the exploitation of stereo information for building detection in off-nadir VHR optical satellite imagery acquired over dense urban areas. The stereo information includes both stereo-based derived elevation (i.e., the height of an object-space point derived photogrammetrically) and measured disparity/parallax (i.e., the relative change of a pixel location due to the change in the viewing angles within a stereo pair) information. It is an article-based PhD dissertation presented through the following papers.

Paper 1 (Peer reviewed):
Suliman, **A**., & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 1913-1923.

Paper 2 (Peer reviewed):
Suliman, **A**., Zhang, Y., & Al-Tahir, R. (2016). Registration-based mapping of aboveground disparities (RMAD) for building detection in off-nadir VHR stereo satellite imagery. *Photogrammetric Engineering and Remote Sensing, 82*(7), 535-546.

Paper 3 (Under review):
Suliman, **A**., & Zhang, Y. (2016). Double projection planes method for generating enriched disparity maps from multi-view stereo satellite images. *The Photogrammetric Record*, (Under Review).

Paper 4 (Under review):
Suliman, **A**., & Zhang, Y. (2016). Disparity-based generation of line-of-sight DSM for image-elevation co-registration to support building detection in off-nadir VHR satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, (Under Review).

1.1 Dissertation Structure

This article-based dissertation includes six chapters. **Chapter 1** provides the introduction to the research. The next four chapters (**Chapter 2** to **Chapter 5**) present the four peer reviewed journal papers listed above, which are either published or under review. In each of the four papers, the first author conducted the primary research, while the second author provided advice on the structure and the remaining authors provided minor input and assistance. **Chapter 6** provides the summary and conclusion of this research. **Figure 1.1** illustrates the organization of this dissertation.

```
┌─────────────────────────────────────────┐
│              (Chapter 1)                 │
│  Dissertation Structure, Background,      │
│  Selected Research Topic, Problem         │
│  Statement, Research Objectives,          │
│  Overview of Each Chapter                 │
└─────────────────────────────────────────┘
```

(**Chapter 2**, Paper 1)	(**Chapter 3**, Paper 2)	(**Chapter 4**, Paper 3)	(**Chapter 5**, Paper 4)
Developing of line-of-sight digital surface model to resolve the image-elevation co-registration challenge	**Registration-based mapping of above-ground disparities** to resolve the disparity gaps and normalization challenge	**Generation of supplementary disparity data** to resolve the disparity gaps and outliers challenges	**Disparity-based elevation generation** of the line-of-sight digital surface model: A combined and improved approach

```
┌─────────────────────────────┐
│        (Chapter 5)           │
│   Summary and Conclusion     │
└─────────────────────────────┘
```

Figure 1.1 Structure of the dissertation

1.2   Background

When the IKONOS satellite was launched in 1999, a new era of earth observation was begun with metre-level ground resolution satellite imagery. This very high resolution (VHR) opened the door for a wide range of earth observation applications. Consequently, the topic of image analysis and information extraction from VHR satellite images has strongly attracted research interest within the remote sensing community.

The VHR optical satellite imagery is normally provided with spatial resolutions of better than 1m for the panchromatic band and up to four metres for the multispectral bands. With the advancements in the remote sensing technologies, the resolution of VHR satellite images has been continuously increasing. Moreover, due to the improved agility of the satellite sensors, these images can be captured in various acquisition modes including corridor, target, stereo, and tri-stereo modes.

As detailed in **Table 1.1**, over 10 VHR optical satellite sensors are currently orbiting the Earth. All of these sensors have off-nadir acquisition capability, and hence they are able to acquire stereo, tri-stereo, and even multi-view stereo VHR images along the same pass (inline/along-track). For privacy protection purposes, the resolutions of the VHR space-borne images were permitted up to 0.5m for panchromatic bands. However, as shown in **Table 1.1**, the highest available resolution of VHR satellite images up to date is 0.31m. This increase was made after relaxing the resolution restriction up to 0.25m by the US Department of Commerce in June, 2014 (DigitalGlobe, 2014). Hence, additional VHR satellites with increased VHR image resolutions are expected to be introduced in the near future.

Table 1.1 Very High Resolutions (VHR) Optical Satellite Sensors[1]

Satellite Sensor	Launch date	Country	Resolution (m)		No. of MS Bands	Swath width (Km)	Max. off-nadir view angle (deg.)
			PAN.	MS.			
OrbView 3	Jun. 2003	USA	1	4	4	8	±50
EROS-B	Apr. 2006	Israel	0.7	N/A	0	7	±45
WorldView-1	Sep. 2007	USA	0.5	N/A	0	17.6	±45
GeoEye-1	Sep. 2008	USA	0.46	1.84	4	15.2	±35
WorldView-2	Oct. 2009	USA	0.46	1.84	8	16.4	±45
Pleiades-1A	Dec. 2011	France	0.5	2	4	20	±47
Pleiades-1B	Dec. 2012	France	0.5	2	4	20	±47
KOMPSAT-3	May.2012	S. Korea	0.7	2.8	4	16	±30
WorldView-3[‡]	Aug.2014	USA	0.31	1.24	8	13.1	±30
TripleSat	Jul. 2015	China	1	4	4	23.4	±45
WorldView-4	Sep. 2016	USA	0.34	1.36	4	14.5	±60

PAN.: panchromatic, MS.: multispectral. [‡] WorldView-3 provides four different band types. It provides one panchromatic band of 31cm resolution, eight multispectral bands of 1.24m resolution, eight short wave infrared of 3.7m resolution, and twelve CAVIS of 30m resolution.

Urban areas are one of the most rapidly and continuously changing land-covers on the Earth. Since buildings are the most prominent objects in urban areas, continuously updated building information, in geographic information systems (GIS), is important for city planning and management applications (Hussain & Shan, 2016). The best geo-spatial data for mapping building information are the VHR satellite images due to their sub-metre resolutions. These resolutions provide extraordinary content of information necessary for mapping complex environments. As a result, building detection in VHR satellite images has become an active research topic in the remote sensing community due to the greater availability, wide coverage, and low cost of the VHR satellite imagery.

The maximum off-nadir viewing angles of the currently working VHR satellite sensors, as indicated in **Table 1.1**, are up to ±60 degrees. Due to this off-nadir acquisition

---

[1] Modified from the technical satellites' specifications, http://www.satimagingcorp.com/satellite-sensors/

capability, most of the archived VHR images are captured off-nadir with along-track and/or across track angles–as can be easily observed in the available online archives of VHR satellite images (e.g., DigitalGlobe archive; *ImageFinder*)–to provide rapid acquisition. Additionally, inline off-nadir image acquisition is usually used to acquire overlapped images that allow stereo information to be extracted. However, although most of the available VHR images are acquired off-nadir, these images are usually avoided in the applications of building detection in dense urban areas as noted in the relevant research publications. This is because off-nadir VHR images captured over urban areas suffer from severe building lean which poses a difficulty for the building detection and mapping methods. Accordingly, this reduces the use and benefits of the off-nadir VHR satellite images.

Building detection methods in VHR satellite images exploit various types of information provided in the images. These information types include textural, spectral, contextual, and morphological image information (Salehi, Zhang, Zhong, & Dey, 2012). Recently, Cheng and Han (2016) conducted a comprehensive survey on the available object detection methods in remote sensing images. They categorized these methods into four groups: template matching-based, machine learning-based, knowledge-based, and OBIA-based (Object-Based Image Analysis) object detection methods. However, since these methods rely on image-based information that lacks the geometric information of the imaged 3D world surface, these methods usually have a limited ability to cope with the inter-similarity among different impervious surfaces in VHR images. For instance, it is very difficult, based on image information only, to automatically distinguish building

roofs from parking lots with the same spectral and spatial properties (Ghaffarian & Ghaffarian, 2014). Therefore, an external type of information is needed.

Considering that buildings are elevated impervious surfaces, elevation information provides an additional key feature for accurate detection and reliable differentiation of building objects from other similar traffic areas. Additionally, as found by Heinl, Walde, Tappeiner, and Tappeiner (2009), the accuracy of urban classification and mapping is increased whenever the elevation data is incorporated regardless of the implemented technique.

Elevation information is commonly generated from different sources of remote sensing data including stereo images and LiDAR (Light Detection And Ranging) data. Among these remote sensing sources, the readily available stereo off-nadir VHR satellite images allow the generation of elevation data to support building detection at a lower cost than other sources of elevations including LiDAR. However, due to the apparent building lean resulting from off-nadir image acquisition, the incorporation of stereo-based information with the off-nadir VHR satellite images for the purpose of building detection still encounters a few challenges and difficulties.

Typically, the incorporation of stereo-based information includes three steps: stereo information generation, co-registration, and normalization. Two types of stereo information can be generated from stereo images: digital elevation models and disparity maps. Hence, the third dimension used in this research for 3D-based detection may be represented by either type of information. Both of the stereo information types need to be accurately co-registered with image data to support identifying the elevated objects. However, since the stereo information is normally generated for the imaged visible

6

surface (i.e., surface stereo information), the effect of the terrain relief is included. Consequently, two equally elevated objects above the ground may have different values of stereo information. To make the detection and delineation of the aboveground (off-terrain) objects directly applicable by thresholding the stereo information, the generated stereo information should be normalized. The normalization process aims to remove the terrain influence by executing two subsequent steps: extracting the terrain effect, and then subtracting it from the generated surface stereo information (Zhang, Tao, & Mercer, 2004). In building detection applications, each of the three incorporation steps introduces some challenges especially when off-nadir VHR satellite images are employed.

1.3 Selected Research Topic

Based on the above mentioned background information, the research topic selected for this dissertation focuses on the application of 3D-based building detection in off-nadir VHR satellite imagery over dense urban areas using stereo information. Central to this topic, the third dimension is the key feature employed for the building detection. This dimension represents the stereo-based information that includes the derived elevation and the measured disparity data extracted from off-nadir VHR stereo satellite images. Hence, the scope of the challenge addressed in this research is the incorporation of the extracted stereo information into a 3D-based method of building detection in satellite images. The problems associated with the off-nadir images as related to the dissertation topic are identified and discussed in the following section.

1.4   Problem Statement

In this research, all of the problems identified for incorporating stereo-based information with off-nadir satellite images result from building lean which is an inherent disadvantage in such an image type. Since off-nadir images are a 2D perspective projection of the 3D real world, the sides of the elevated objects will be captured as leaning objects and hence they add more complexity to the acquired scenes. More specifically, the façades of buildings in off-nadir images cover the area between building roofs and their corresponding footprints. Consequently, the building roofs will be separated from their footprints. This separation creates offsets between the building roofs and their correct orthographic locations.

Both the building roof offsets and building façades create four challenges in the incorporation of stereo information with off-nadir images for the purpose of mapping building rooftops. On the one hand, the offsets of the building roofs prevent efficient image-to-elevation co-registration. Additionally, these offsets vary based on the building heights which are not constant. As a result of that, the generated detection maps of building roofs in off-nadir images will be prevented from being directly geo-referenced for easy integration with other existing maps. On the other hand, the building façades create both occluded areas and confusing similarity for the automatic matching algorithms particularly when off-nadir stereo mates are employed. All of these problems associated with off-nadir images, as described in the following subsections, reduce the quality of the 3D-based building detection results.

1.4.1 Image-Elevation Co-registration

The first challenge is related to the multi-source data co-registration. In other words, stereo-based information (i.e., 3D structure information) is required to be co-registered with the corresponding image data to support building detection. The challenge results from the offsets of the roofs in off-nadir images. In the case of disparity information, measured disparity maps created by matching techniques are co-registered perfectly, as per disparity definition, to the selected reference stereo image. Hence, no effort is required to accurately co-register the disparity information even with off-nadir images. In contrast, accurate co-registration of elevation data with off-nadir VHR satellite images acquired over dense urban areas is very challenging (Suliman & Zhang, 2015). This results mainly from the dissimilarity of the projections of the two data types being co-registered. While the 3D elevation models typically have an orthographic projection, the 2D images have perspective projections. In the elevation data, consequently, the building roofs will occupy the same locations as their corresponding footprints. However, this is not the case for off-nadir images because the building roofs are shifted from their corresponding footprints.

The most successful technique that provides accurate elevation co-registration even with off-nadir images is true orthorectification (Guoqing, Weirong, Kelmelis, & Deyan, 2005; Chen, Teo, Wen, & Rau, 2007; Habib, Kim, & Kim, 2007). However, the implementation of this technique is expensive (in terms of input data and computation) and the orthoimages are difficult to achieve. Hence, the first concern of this dissertation is to address the problem of efficient image-elevation co-registration for off-nadir VHR satellite images. The proposed solution should provide pixel-level co-registration

accuracy and reduce the required implementation cost (e.g., input data, calculation steps, and computation cost) in order to be considered as a more efficient technique than what is currently available.

1.4.2   Disparity Gaps and Normalization

The second challenge associated with off-nadir VHR satellite images is the severe occlusion effects resulting from building façades in off-nadir stereo images acquired over dense urban areas. Since these images are 2D perspective projection of the 3D real world, the elevated objects, mainly the building façades, will hide large areas behind. These occluded areas will appear only in one of the stereo images and thus no successful matching could be made. As a result, many data gaps will be created in the measured disparity maps (i.e., disparity gaps).

Filling the resulting disparity gaps by surface interpolation techniques, as the typical solution for the case of one stereo pair, in dense urban areas reduces the quality of the generated maps by producing misleading surface information. This wrong information affects negatively the subsequent process of normalizing the surface stereo-based information which is commonly required to remove the terrain influence and have only the stereo-based information representing the off-terrain objects. Since the stereo-based information typically represents the surface of the visible objects which include the terrain and off-terrain level ones, the normalization process includes two steps: extracting the terrain-level stereo information and then subtracting it form the generated surface information. As a result of that, if the filled gaps in the generated disparity maps have the interpolation-based misleading surface information, the results of extracting and

subtracting the terrain-level stereo information from that disparity maps will be also inaccurate.

Therefore, the second concern of this dissertation is to mitigate the problem of the disparity gaps and the process of stereo-based information normalization by measuring directly the aboveground (i.e., normalized) disparity information to bypass the steps of interpolating the occlusion disparity gaps and removing the terrain relief influence.

1.4.3   Disparity Gaps and Outliers

The third difficulty of off-nadir VHR satellite images is associated with the effect of both occlusion and confusing similarity when stereo off-nadir images are utilized to generate the disparity information.

In off-nadir satellite images over urban areas, building façades usually hide large areas in one image of the stereo pair. The matching results in such occluded areas are always unsuccessful. These unsuccessful matches are either failed matches (creating gaps) or false matches (creating outliers). Moreover, different building façades may have a similar appearance that confuses matching algorithms. Hence, additional false matches are created. For detecting these false matches, visibility checks (e.g., the left-Right check) and/or consistency checks (i.e., comparing different disparity values from different stereo-pairs for the same object-space point) are typically used. Once the false matches are detected, the corresponding incorrect (or unstable) disparity values are then removed. Afterward, the resulting disparity gaps from removing the outliers are then replaced with the correct values. Therefore, both failed and false matches produce disparity gaps that

need to be filled correctly. Thus, disparity maps generated from stereo off-nadir images over dense urban areas usually suffer from large data gaps.

The resulting gaps are typically filled either from the surrounding pixels (i.e., as generated by using the interpolation techniques or 2D convolution filters) or from additional external data. The techniques that rely on the data from the surrounding pixels perform poorly in the case of large data gaps. Hence, the techniques that rely on generating additional data are preferred.

In the case of multiple stereo images, extra stereo-based elevation data can be easily generated and directly applied for both detecting the outliers and filling the gaps since the derived elevations for the same ground point are equal regardless of the stereo pairs used. However, this advantage of generating supplementary data is not directly available/ possible for the disparity information measured from different stereo pairs. This is because the disparity value –which is the relative location shift of a shared point within a stereo pair –is a relative co-relation between two stereo images. Thus, the third concern of this dissertation is to address the problem of generating supplementary disparity data from different inline stereo pairs that are applicable for enriching disparity maps by filling gaps and correcting outliers.

1.4.4 Direct Map Geo-referencing

The fourth and last problem connected with off-nadir VHR satellite images is related to the challenge of building map geo-referencing. The offset between a building rooftop and its footprint is created due to the off-nadir acquisition mode. Furthermore, this offset varies from one building to another based on the elevation of each building individually.

As a result, the detection building maps cannot be directly geo-referenced for easy integration with other geo-referenced maps and/or GIS layers since the rooftop-to-footprint offsets are variable. Any building detection method that uses image-based information and cues cannot shift the detected rooftops to their correct corresponding footprints unless the elevation information is available.

Hence, to close the cycle of building detection by map geo-referencing, this dissertation addresses the problem of correcting the roof offsets based on deriving ground elevations efficiently from disparity information to allow geo-referencing the mapped roofs in off-nadir VHR stereo satellite images. By constructing a direct disparity-to-elevation relationship, supplementary disparity and elevation data can be efficiently generated, and accordingly the roof offsets can be corrected. Hence, these problems should be addressed in the same context since they are co-related.

1.5   Research Objectives

The objectives of this research are fourfold in order to solve/address the four identified problems in a progressively improving manner mainly in terms of effort and computational steps. All of the identified problems are meant to be addressed in the context of 3D-based building detection in off-nadir VHR stereo satellite images acquired over urban areas to demonstrate the applicability of the developed solution. The urban areas in this research are characterized by moderate terrain relief variation with dense and highly elevated buildings. Therefore, the four main objectives are described in the following subsections.

1.5.1   Efficient Image-Elevation Co-registration

Co-registering the image data to the relevant elevation information is critical for all elevation-based building detection methods. Based on a review of the relevant literature, three techniques were found to be able to provide image-elevation co-registration: 2D-based registration, orthorectification, and true orthorectification. For the case of employing off-nadir VHR images over dense urban areas, none of these techniques can be considered efficient in terms of both accuracy of co-registration, and cost of data and computation. Hence, the first objective of this dissertation, which is addressed in **Chapter 2**, is to develop an accurate solution for image-elevation co-registration with inexpensive implementation cost in input data, special algorithms, and computational steps. The ultimate goal of this objective is to demonstrate the applicability of the proposed co-registration technique in elevation-based building detection in off-nadir VHR satellite imagery acquired over dense urban areas.

1.5.2   Mapping of Aboveground Disparities

Normalizing the stereo information that is co-registered to the relevant imagery is an important step to support building detection in VHR images. The use of disparity instead of stereo-based elevation information improves the efficiency of the building detection method by both providing pixel-level accuracy of co-registration and reducing the steps of computation. Additionally, the number of input images would be minimized to just two images since the disparity maps are typically generated from a pair of stereo images. However, in the case of using stereo pairs of off-nadir VHR satellite images, large data gaps in the measured disparity maps are produced due to the severe façade occlusion in

such images. These gaps create misleading information in the normalization process that affects the quality of the 3D-based building detection. Therefore, the second objective of this dissertation, which is addressed in **Chapter 3**, is to develop a technique for mapping the aboveground disparities to bypass the typical processing steps of gaps interpolation and disparity normalization that both create misleading information. The ultimate goal of this objective is to demonstrate the applicability of the proposed mapping technique in disparity-based building detection in off-nadir VHR satellite imagery acquired over dense urban areas.

1.5.3 Supplementary Disparity Data Generation

Stereo information generated from a pair of off-nadir images usually suffers from occlusion-produced gaps. Filling the data gaps by interpolation techniques usually creates misleading surface information because these techniques are localized. However, if supplementary disparity information is available, superior enhancement (by filling the gaps based on external data source) should be expected.

Since corresponding disparity values from different stereo pairs are usually inconsistent, co-relating the relevant disparities to their corresponding object-space elevation should build proportionality and eliminate the disparity inconsistency. Therefore, the third objective of this dissertation, which is addressed in **Chapter 4**, is to construct scale-related (inter-proportional) disparity in the object space among different stereo pairs for generating supplementary information from multi-view inline stereo VHR satellite images for disparity map enhancement. The enhanced disparity maps should improve the quality of the building detection in VHR satellite images.

1.5.4   Disparity-based Elevation Generation and Map Geo-referencing

Generating elevations based on a proportionality relationship with disparity values allows both efficient elevation generation from disparity maps and direct co-registration with optical images. Based on this disparity-elevation proportionality, the main objective of **Chapter 5** is to achieve efficient image-elevation co-registration through disparity-based elevation generation. Therefore, relevant to this objective is to demonstrate the applicability of the generated and co-registered elevation data in elevation-based building detection in off-nadir VHR satellite imagery acquired over a dense urban area.

Geo-referencing the detected building roofs is essential for map integration with other existing maps and GIS layers. Since the off-nadir acquisition of the VHR satellite images variably offsets the building roofs from their corresponding footprints, the detection map generated for the building roofs could not be geo-referenced directly by the conventional techniques. The roof offsets are varying depending on the elevation of each building individually. Hence, to correct the roof offsets and to allow direct geo-referencing, elevation information is required for each building for shifting the roofs' images to their correct footprint locations. Thus, the subsequent and last objective of **Chapter 5** is to correct, based on the derived elevation data, the building-roof offsets in off-nadir images and geo-reference the detection map.

In relation to the dissertation structure, the four identified research objectives are carried out in four chapters (**Chapter 2-5**). While the first three objectives are discussed separately in **Chapters 2**, **3**, and **4** respectively, the fourth objective is addressed along with its subsequent one in **Chapter 5**.

1.6   Data and Metrics

The data and metrics used to evaluate the methods developed in each chapter are

summarized in **Table 1.2**.

Table 1.2 Data and metrics used for evaluating the proposed methods.

Source	Test Data	Metric Description	Chapter
Airbus Defence and Space (Sample Data)	A subset of tri-stereo VHR satellite images acquired by the Pleiades-1A sensor over the urban area of Melbourne, Australia on February/2012. The off-nadir acquisition angles of the forward and backward images are approximately $+15^{\circ}$ and -15° respectively. The stereo pairs are used to derive photogrammetrically digital surface model elevations to be co-registered with one of the off-nadir images.	*Quantitative assessment:* Correctness, Completeness, and Overall Quality performance measures of building detection methods (Rutzinger, Rottensteiner, & Pfeifer, 2009).	**Chapter 2** (Suliman & Zhang, 2015)
Prof. C. Fraser Univ. of Melbourne	A set of ground control points (GCPs), of accuracy within 10-15cm, were used to validate the quality of the stereo-based derived data.		
Airbus Defence and Space (Sample Data)	A subset of a stereo pair of VHR satellite images acquired by the Pleiades-1A sensor over the urban area of Melbourne, Australia on February/2012. The two opposite off-nadir angles of the forward and backward images are approximately $+15^{\circ}$ and -15° respectively.	*Quantitative assessment:* Correctness, Completeness, and Overall Quality performance measures of building detection methods (Rutzinger et al., 2009).	**Chapter 3** (Suliman, Zhang, & Al-Tahir, 2016)
Airbus Defence and Space (Sample Data)	A subset of tri-stereo VHR satellite images acquired by the Pleiades-1A sensor over the urban area of Melbourne, Australia on February/2012.	*Qualitative assessment:* Isometric/3D rendered views *(Congalton & Green, 2009)*	**Chapter 4** (Suliman & Zhang, 2016b)
Prof. C. Fraser Univ. of Melbourne	A set of GCPs (10-15cm accuracy) were used to validate the quality of the stereo-based derived data. These data are then used to evaluate the quality of the data derived from an enriched disparity map.	*Quantitative assessment:* Root-mean-square error (RMSE).	
DigitalGlobe via IGARSS-2011 (Contest Data)	A subset of five stereo VHR satellite images acquired by the WorldView-2 sensor over the urban area of Rio de Janeiro, Brazil on January/2010. The off-nadir acquisition angles of these images are approximately $+45^{\circ}, +33^{\circ}, +8^{\circ}, -30^{\circ},$ and -45°	*Quantitative assessment:* Correctness, Completeness, and Overall Quality performance measures of building detection methods (Rutzinger et al., 2009).	**Chapter 5** (Suliman & Zhang, 2016a)

1.7   Overview of Each Chapter

   Chapter 1 is the introduction. It comprises the structure of the dissertation, research background and topic selection, problem statement, objectives of the research, and an overview of each remaining chapter.

Chapters 2 to 5 contain the four journal papers representing the main contributions of this PhD dissertation.

- **Chapter 2** introduces the line-of-sight digital surface model solution to co-register efficiently photogrammetric-based elevation data even with off-nadir VHR satellite images with some advantages over the currently available methods. It represents a critical step for supporting the methods of elevation-based building detection in off-nadir VHR satellite images using photogrammetric-based elevation data.

- **Chapter 3** represents an efficiency improvement over the use of elevation data in the required incorporation steps by utilizing disparity data. This chapter demonstrates a new registration-based method for mapping the aboveground disparities of off-terrain features in reasonably non-flat dense urban areas. It bypasses the typical normalization steps for supporting disparity-based building detection in off-nadir VHR satellite images using one stereo pair of off-nadir images.

- **Chapter 4** represents a double projection planes method for analytically building a proportionality relationship among the disparity values from different stereo pairs of multi-view VHR satellite images. The method computes disparity transferring scales that are utilized to generate supplementary disparity data to

allow generating enriched disparity maps by filling the occlusion disparity gaps in the measured maps.

- **Chapter 5** represents a combination of the achieved advantages in the three previous chapters to generate and enhance efficiently, using supplementary data, a disparity-based line-of-sight digital surface model that is co-registered to an off-nadir VHR satellite image with pixel-level accuracy. Additionally, the chapter demonstrates the applicability of the generated stereo information for detecting building roofs in off-nadir VHR satellite images and geo-referencing the mapped roofs successfully.

Chapter 6 presents the conclusions. It summarizes the achievements of this research and outlines its drawbacks and limitations. It also presents some recommendations for future research.

In relation to the dissertation structure, the developed chapters are connected through a progressive improvement manner in the required calculation steps. While **Chapter 2** attempts to improve the efficiency of the currently available co-registration techniques, **Chapter 3** boosts the improvement in the computation cost over **Chapter 2** for the dense and moderately non-flat urban areas by using disparity instead of elevation data. **Chapter 4** attempts to produce enriched disparity maps through generating supplementary disparity data. The concept of this chapter is extended in **Chapter 5** to derive elevation data from disparity maps. Hence, **Chapter 5** represents a combined improvement over the previous three chapters mainly by keeping the achieved advantages and reducing the computational steps of the algorithm developed in **Chapter 2**.

REFERENCES

Chen, L., Teo, T. , Wen, J. , & Rau, J. (2007). Occlusion compensated true orthorectification for high-resolution satellite images. *Photogrammetric Record, 22*(117), 39-52.

Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *117*(7),11-28. doi:10.1016/j.isprsjprs.2016.03.014

Congalton, R., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton: CRC Press/Taylor & Francis.

DigitalGlobe. (2014). U.S. department of commerce relaxes resolution restrictions digitalglobe extends lead in image quality. Retrieved from digitalglobe investor Resolutions: http://investor.digitalglobe.com/

Ghaffarian, S., & Ghaffarian, S. (2014). Automatic building detection based on supervised classification using high resolution google earth images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(3), 101–106.

Guoqing, Zhou, Weirong, Chen, Kelmelis, J. A., & Deyan, Zhang. (2005). A comprehensive study on urban true orthorectification. *IEEE Transactions on Geoscience and Remote Sensing, 43*(9), 2138-2147.

Habib, A., Kim, E., & Kim, C. (2007). New methodologies for true orthophoto generation. *Photogrammetric Engineering and Remote Sensing, 73*(1), 25-36.

Heinl, M., Walde, J., Tappeiner, G., & Tappeiner, U. (2009). Classifiers vs. input variables—the drivers in image classification for land cover mapping. *International Journal of Applied Earth Observation and Geoinformation, 11*(6), 423-430.

Hussain, E., & Shan, J. (2016). Urban building extraction through object-based image classification assisted by digital surface model and zoning map. *International Journal of Image and Data Fusion, 7*(1), 63-82. doi: 10.1080/19479832.2015.1119206

Rutzinger, M., Rottensteiner, F., & Pfeifer, N. (2009). A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2*(1), 11-20.

Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of Geoinformatics, 8*(2), 35-51.

Suliman, A., & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 1913-1923.

Suliman, A., & Zhang, Y. (2016a). Disparity-based generation of line-of-sight DSM for image-elevation co-registration to support building detection in off-nadir VHR satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* (Under Review).

Suliman, A., & Zhang, Y. (2016b). Double projection planes method for generating enriched disparity maps from multi-view stereo satellite images. *The photogrammetric Record,*(Under Review).

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016). Registration-based mapping of aboveground disparities (RMAD) for building detection in off-nadir VHR stereo satellite imagery *Photogrammetric Engineering & Remote Sensing, 82*(7), 535-546.

Zhang, Y., Tao, C., & Mercer, J. (2004). An initial study on automatic reconstruction of ground DEMs from airborne IfSAR DSMs. *Photogrammetric Engineering and Remote Sensing, 70*(4), 427-438.

Chapter 2: DEVELOPMENT OF LINE-OF-SIGHT DIGITAL SURFACE MODEL FOR CO-REGISTERING OFF-NADIR VHR SATELLITE IMAGERY WITH ELEVATION DATA[1]

Abstract

Co-registration of very high resolution (VHR) images with elevation data is extremely important for many remote sensing applications due to the complementary properties of these two data types. However, this type of multi-data source registration has many associated challenges. For instance, although VHR satellite images are mostly acquired off-nadir, the integration of off-nadir images with digital surface models (DSM) for the purpose of urban mapping has been rarely seen in research publications. This is due to the relief displacement of the elevated objects which causes a problematic misregistration between the perspective off-nadir images and the corresponding orthographic DSMs. Therefore, the co-registration of such datasets is almost impossible unless a true orthorectification process is executed. However, true orthoimages are expensive, time consuming, and difficult to achieve. Thus, this paper proposes a registration method based on developing a line-of-sight DSM solution to effectively register elevation data

---

with off-nadir VHR images. The method utilizes the relevant sensor model in two phases: (1) deriving DSM from stereo images, and (2) re-projecting the DSM back to one of the stereo images to generate a line-of-sight DSM for accurate co-registration. To demonstrate the applicability of the proposed method and evaluate the effect of the misregistration, a building detection procedure is implemented. The proposed method is found to be feasible, inexpensive and of sub-pixel accuracy. Additionally, it improves the overall accuracy of detecting buildings by almost 12% relative to that when the conventional 2D registration technique is used solely due to the elimination of the misregistration effect.

2.1   Introduction

The accurate integration or registration of optical images with elevation datasets (Optical-elevation data registration) plays an important role in many remote sensing applications. Examples of such applications include image analysis and classification (Salehi, Zhang, & Zhong, 2011), building detection and extraction (Chen, Zhao, Han, & Li, 2012), 3D city modelling (Fujii & Arikawa, 2002), and canopy modeling (Chen, Chiang, & Teo, 2005). The mapping products of these applications, and the remote sensing images, are the main, and sometimes the only, database inputs for Geo-Informatics Systems (GIS) (Salehi, Zhang, Zhong, & Dey, 2012).

The modern advancements in remote sensing technology have led to the availability of optical very high resolution (VHR) images. These VHR images are characterized by rich information content and texture, broad coverage, fast acquisition and a relatively low price (Mishra & Zhang, 2012). Due to the need for fast and various acquisition modes,

however, most of the time VHR sensors capture images with some across and/or along-track angle producing off-nadir images (Hong & Zhang, 2008). These images always have an apparent relief displacement effect when they are acquired over urban areas. Furthermore, optical images are the 2D perspective projection of the 3D world; therefore, height information about the real surface is missed.

Elevation data, which are typically in the form of digital elevation models (DEMs), provide the height information about the real world surface. For instance, while digital terrain models (DTMs) describe terrain elevations, digital surface models (DSMs) describe the visible surface, which includes the elevations of both terrain and off-terrain objects (Sefercik, Karakis, Bayik, Alkan, & Yastikli, 2014). When these two elevation models are subtracted from each other, the so-called normalized digital surface model (nDSM) is calculated (Weidner & Förstner, 1995). The common sources for generating elevation data are photogrammetric approaches and Light-Detection-And-Ranging (LiDAR) technology. However, the produced elevation models, unlike optical images, lack the spectral, radiometric and texture information.

The optical imagery and elevation data have complementary properties. Therefore, the integration of these two datasets is a valid approach to make one data source that compensates for their limitations. Consequently, this integration will lead to higher quality surface information and provide reliable key features for extracting information from remotely sensed images (Mishra & Zhang, 2012). The most successful and recent image-based building detection algorithms from VHR satellite imagery, such as those presented and evaluated by Khosravi, Momeni, and Rahnemoonfar (2014) and those by Stankov and He (2014), cannot reliably delineate different impervious classes with inter

24

similarity image information (i.e., spectral, morphological, textural, and contextual information). For instance, building rooftops and parking lots are impervious classes that are likely to have the same material and shape properties. Thus, when such a case exists in a satellite scene, both pixel and object-based classification or detection techniques, those relying on the image information only, fail to reliably delineate such spectrally-spatially similar classes. However, by integrating ancillary data (e.g., nDSM or GIS layers), a more reliable delineation can be achieved. This is due to incorporating key delineation components that are independent from the image information such as elevations.

In addition to the reliability, the accuracy of classification and building detection is increased when the elevation data are incorporated regardless of the implemented technique (Heinl, Walde, Tappeiner, & Tappeiner, 2009; Salehi et al., 2012). However, several problems are introduced when multi-data sources are integrated, particularly the problematic misregistration with the ancillary data when off-nadir VHR images are employed (Dowman, 2004; Salehi et al., 2012).

Therefore, the objectives of this study are (1) to address the problem of how to accurately co-register elevation data with off-nadir VHR optical images acquired over urban areas, (2) to propose an optical-elevation data registration method along with an assessment procedure, (3) to develop a program tool to implement the proposed method, and (4) to demonstrate the applicability of the proposed method and evaluate its performance in a building detection application using ancillary elevation data (i.e., stereo-based DSM).

Consequently, the contribution of this paper is the development of a line-of-sight DSM solution for the purpose of sub-pixel co-registration between optical and elevation data even for off-nadir VHR imagery (i.e., regardless of the acquisition angle) acquired over dense urban areas. This solution solves an important problem in urban land cover classification: because most VHR images are off-nadir (due to the agile point-and-shoot image collection of VHR satellites) and conventional co-registration techniques cannot successfully register elevated objects with the corresponding DSM, significant classification errors arise. To the best of the authors' knowledge, this solution is first proposed by this paper, a statement supported by an extensive review of the literature. The current study represents a significant extension of a preliminary work presented briefly in Suliman and Zhang (2014).

2.2   Review of Existing Methods

Based on the reviewed literature, three approaches have been proposed for the purpose of optical-elevation data integration: (1) image-to-image registration, (2) orthorectification, and (3) true-orthorectification. Unlike the case of off-nadir images, all of these approaches provide high quality results when nadir images are employed.

Image-to-image registration is a mapping process between two images both spatially and with respect to intensity (Brown, 1992; Xiong & Zhang, 2010). It aims to geometrically align one image with another one of the same scene which may differ in viewpoints, times, or sensors (Fonseca & Manjunath, 1996; Habib & Al-Ruzouq, 2005). Thus, the registration process is a critical phase in some applications where the final information is extracted from the integration of different data sources. Image registration

fusing optical and LiDAR datasets is used successfully both for forest canopy modeling as presented by Chen et al. (2005) and for 3D urban scene representation as demonstrated by Mastin, Kepner, and Fisher (2009). The achieved results were of high quality; however, the employed images were restricted to be of nadir views. Comprehensive surveys and reviews of image registration methods are provided by Brown (1992), Zitová and Flusser (2003), and Xiong and Zhang (2010). In particular, the existing registration methods of optical and LiDAR datasets were reviewed by Mishra and Zhang (2012).

Based on these reviews, it can be concluded that accurate image registration of optical and elevation datasets is difficult to achieve due to the substantial differences between these two datasets (Palenichka & Zaremba, 2010). For example, while off-nadir VHR images have perspective viewing that causes a highly apparent relief displacement over urban areas, elevation data have orthographic projection (nadir viewing). This substantial difference in the projections makes the pixel by pixel registration almost impossible (Salehi et al., 2011). Thus, the incorporation of sensor information for orthorectification process is strongly recommended by many researchers to eliminate the relief displacement of the optical images (Hong & Zhang, 2008; Mishra & Zhang, 2012).

The orthorectification process aims to eliminate the perspective and relief displacement distortions from the acquired optical imagery by re-projecting it to the object space using the relevant sensor and elevation models. As a result, the produced orthoimages will have a photographic appearance, uniform scale, and orthographic projection (Habib, Kim, & Kim, 2007). Additionally, they will be geo-referenced to the ground coordinate system of the employed height model. However, traditional orthorectification algorithms rely solely on DTMs where the off-terrain elevations are not

considered. Therefore, such algorithms fail to orthorectify off-terrain objects, like high-rise buildings, particularly those of apparent lean in off-nadir VHR images. Consequently, true-orthorectification is required.

True-orthorectification approaches, unlike the traditional orthorectification ones, use high quality DSMs and consequently implement algorithms for occlusion detection and compensation for the hidden areas to produce truly accurate orthographic images (Sheng, Gong, & Biging, 2003). However, these approaches are computationally expensive and time consuming, and require multiple overlapped images to fill the occluded areas (Chen, Teo, Wen, & Rau, 2007). Moreover, the radiometric differences among the input overlapped images may degrade the orthoimage visual quality in the compensation and resampling process (Karras, Grammatikopoulos, Kalisperakis, & Petsa, 2007).

As evident from this review, pixel by pixel co-registration of perspective off-nadir VHR optical images with orthographic elevation data is very challenging and hard to achieve. Therefore, none of the reviewed approaches can be considered as an efficient way of optical-elevation data registration for off-nadir VHR images when they are captured over urban environments. The real challenge of such registration worsens when high-rise buildings appear in the off-nadir VHR images due to the extreme relief displacement of their rooftops. Hence, the question that arises in this study is how to register stereo-based elevations with off-nadir VHR optical images acquired over urban areas in a way that mitigates or overcomes most of the above mentioned limitations.

Referring to the reviewed literature, the incorporation of sensor model information is found to be critical for accurate registration. Furthermore, the image resampling process usually degrades radiometric quality and changes the spectral information of the input

imagery. Thus, if the sensor information is used to re-project/transfer the elevation data to the image space and produce what may be called a line-of-sight DSM, accurate registration will be achieved regardless of the acquisition angle of the image used. Additionally, due to avoiding the image resampling process, the original spectral and radiometric information of the employed images will be preserved.

The remainder of the paper is organised as follows: **Section 2.3** describes the proposed line-of-sight DSM registration method and its accuracy assessment procedure. **Section 2.4** demonstrates the applicability of the proposed method in a building detection application along with its performance evaluation. The dataset used and the results achieved, for both registration and building detection, are presented and discussed in **Section 2.5**. Lastly, **Section 2.6** gives the conclusions.

2.3 Methodology

This research work proposes that the registration of the DSMs' elevations with off-nadir VHR optical imagery can be achieved by projecting the ground elevations back to the image space using the relevant sensor model in order to develop a kind of line-of-sight DSM as described in **Section 2.3.1**. Furthermore, the registration assessment, as in **Section 2.3.2**, is adapted to be quantitative and qualitative.

2.3.1 The Proposed Line-of-Sight DSM Registration Method

The proposed registration method is flowcharted in **Figure 2.1**. The figure shows that there are two cases for the availability of the elevation data: pre-existing DSM (path-B) and generated DSM (path-A). While the former includes the already available DSMs,

the latter includes only the case when the optical image being registered is used with its stereo mates to photogrammetrically generate elevation data.
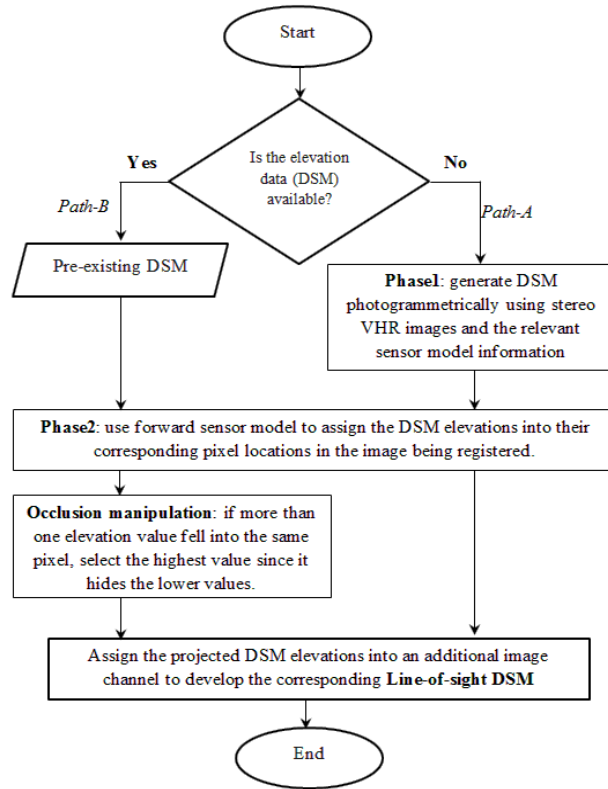
Figure 2.1 The flowchart of the proposed line-of-sight DSM registration method for co-registering optical with elevation data.

Since the LiDAR-derived DSMs are expensive and not available for most places, it is proposed in this study for the height information to be derived photogrammetrically from VHR stereo images either of which can be used for the registration. Therefore, the implemented method in this paper consists of two phases (Phase 1 and Phase 2) as shown in **Figure 2.1** path-A.

The two phases of the implemented method are detailed in **Figure 2.2**. It is clear, from this figure, that sensor model information is essential. In Phase 1 a DSM is

generated from stereo VHR satellite images using the backward/inverse sensor model. However, in Phase 2 the generated DSM elevations are re-projected from ground space back to the image space by using the forward sensor model information. Therefore, each image pixel will have its corresponding ground height. This is true for all pixels except those of low textured or occluded areas due to the absence of the elevation data. The input imagery will have an additional layer that contains the assigned and registered DSM ground elevations (i.e., line-of-sight DSM).

Phase 1: stereo-based elevation data generation using inverse sensor model

Phase 2: Back projection of the elevations using forward sensor model
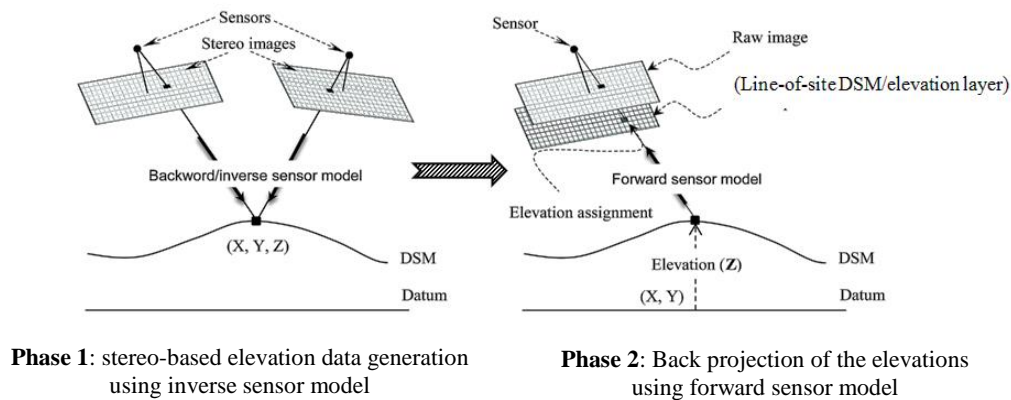
Figure 2.2 A more detailed illustration for the two phases of the proposed registration method.

When a pre-existing DSM is employed, more than one elevation value might compete for the same pixel due to occlusion effect (see **Figure 2.3**). Thus, to handle this effect as in the manipulation step of **Figure 2.1**, a perspective DSM (i.e., image space elevation layer) would be generated by assigning the highest elevation values since such values occlude/hide the lower ones as illustrated in **Figure 2.3**. The generated elevation layer, then, will contain the line-of-sight DSM which represents the perspective projected elevations registered to their correct corresponding locations on the optical imagery.

31

Sensor models, as shown in **Figure 2.2**, are required to establish the relationship between image space and object space. This relationship is crucial for many photogrammetric reductions including elevation data generation and image orthorectification. The sensor models are classified typically into two types: physical (rigorous) and generalized (approximated) sensor models. Unlike the generalized models, physical models, such as Collinearity equations, represent the imaging process through positional and orientation parameters. However, the rational function model (RFM), which is the most popular generalized sensor model, uses a pair of ratios of two polynomials to approximate the Collinearity equations via a set of rational polynomial coefficients (RPCs) that do not have physical meanings (Tao & Hu, 2001).
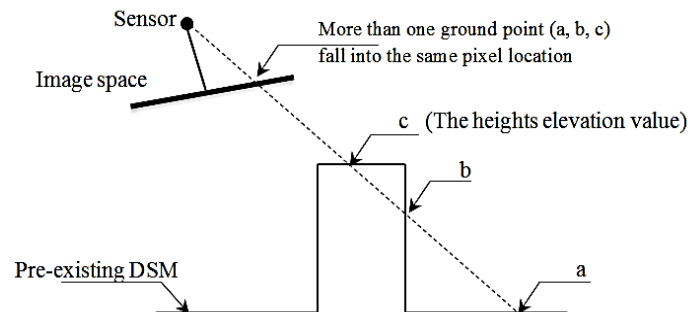
Figure 2.3 Occlusion effect and its manipulation when a pre-existing DSM is used.

For VHR satellite images that have a small field of view and high-precision orientation parameters, the RFM has been successfully applied due to its good approximation in terms of geometrical processing. The orientation parameters of the VHR satellite images, including the orbit parameters and altitude data, are not available for end users; therefore, RPCs are provided instead (Tao & Hu, 2001). Two types of RFM model are available: the forward and the inverse model. Consequently, two sets of

RPCs are usually provided. The forward model provides the mapping from object space to image space, while the inverse model provides the opposite mapping. The mathematical expression for the forward RFM model is provided in the following equations as given in Tao and Hu (2001) and Fraser and Hanley (2003) .

$$L_n = \frac{\sum_{i=0}^{i=3}\sum_{j=0}^{j=3}\sum_{k=0}^{k=3} a_t X_n^i Y_n^j Z_n^k}{\sum_{i=0}^{i=3}\sum_{j=0}^{j=3}\sum_{k=0}^{k=3} b_t X_n^i Y_n^j Z_n^k} \quad , \quad S_n = \frac{\sum_{i=0}^{i=3}\sum_{j=0}^{j=3}\sum_{k=0}^{k=3} c_t X_n^i Y_n^j Z_n^k}{\sum_{i=0}^{i=3}\sum_{j=0}^{j=3}\sum_{k=0}^{k=3} d_t X_n^i Y_n^j Z_n^k} \qquad (2.1)$$

where L_n and S_n are the normalized image coordinates (line and sample). X_n, Y_n, and Z_n are the normalized ground coordinates. The normalization is to standardize the values to be within [-1,+1] through translation/offset and scale as follows:

$$L_n = \frac{L - L_O}{L_S} , S_n = \frac{S - S_O}{S_S} , X_n = \frac{X - X_O}{X_S} , Y_n = \frac{Y - Y_O}{Y_S} , Z_n = \frac{Z - Z_O}{Z_S} \qquad (2.2)$$

where L and S are the de-normalized image coordinates. X, Y, and Z are the de-normalized ground coordinates. X_O, Y_O, Z_O, X_S, Y_S, and Z_S are the offset and scale values for the three ground coordinates. The maximum power of each ground coordinate is limited to 3 (i.e., i, j, and k $\leq$ 3) and the total power of all ground coordinates is limited to 3 (i.e., i+j+k $\leq$ 3). In this case, each polynomial is of 20-term cubic form. a_t , b_t , c_t, and d_t are the 80 coefficients (RPCs) of the rational function model (t = 1,2,3 …20). Usually, RPCs have a systematic bias which can be compensated by an affine transformation. A small number of ground control points (GCPs) are needed for the systematic compensation (Fraser & Hanley, 2003).

In this study, to carry out Phase 2 and assign the ground elevations in their corresponding image locations, the forward RFM model is proposed to be employed.

33

This is because it is easier and faster than the inverse model. It inexpensively calculates the image location, unlike the inverse model, without any iterative process. Furthermore, for security reasons, most of the satellite imagery vendors provide the forward model coefficients (i.e., forward RPCs) to the general users; additionally, it can be used directly when a pre-existing DSM is available.

Thus, the involved three steps of the proposed optical-elevation data registration method are elevation data extraction, elevation data re-projection, and line-of-sight DSM generation. They can be further described as follows:

2.3.1.1 Extracting stereo-based elevation data:

This process photogrammetrically derives elevation data (i.e., DSM) from stereo images by executing three main steps:

(a) *Epipolar image generation* — epipolar images are re-projected stereo pairs so that the left and right images have a common orientation, and matching features between the images appear along a common x-axis. This property extremely increases the efficiency of dense matching process for DEM extraction. A comprehensive analysis of the epipolar geometry of the images acquired by frame and linear array sensors is provided by Habib, Morgan, Jeong, and Kim (2005).

(b) *Dense image matching* — matching points in the left and right input epipolar images are found using image correlation. Area-based pyramid matching approach (Lehner & Gill, 1992) is a well-established approach that is implemented by many photogrammetric software packages for dense matching due to its speed and accuracy.

(c) *Forward intersection*— intersecting the light rays of the matched image points in the stereo images produces the 3D ground locations. The image coordinates of the matched points are combined with the relevant sensor model in a block adjustment approach to extract the geocoded DSM elevation model. The adjustment process of VHR satellite imagery described by RPCs is discussed in Grodecki and Dial (2003).

2.3.1.2   Re-projecting the elevation data:

As described in **Equation (2.1)**, the forward RFM sensor model provides the pixel location of a point as a function in its normalized 3D geographic location. Therefore, by having the RPCs of the forward RFM sensor model, a direct projection of the generated (or the pre-existing) 3D ground coordinates can be achieved. A program tool was developed, for this study, to perform the necessary geographic transformation and to calculate the planimetric pixel location in the image space.

2.3.1.3   Generating the line-of-sight DSM:

After finding the pixel locations of the corresponding ground points, an additional image band is generated to accommodate the projected ground elevations for accurate pixel-by-pixel registration. This step is included in the developed program tool. In the case of occluded areas, as described in **Figure 2.3**, the higher elevation value will be selected.

The proposed steps of this optical-elevation datasets registration seem to be simple and inexpensive, and should facilitate the buildings' hypothesis generation and support

the roof detection of the high buildings in VHR remote sensing images, even in off-nadir ones. The next subsection describes the adapted assessment procedure.

2.3.2   Registration Assessment

To assess the proposed method, the result should be compared against the already available techniques. Since a pre-existing dense and accurate DSM is not available for the study area, true-orthorectification could not be implemented to be compared with the proposed method. Moreover, despite the expected high registration accuracy, true-rectification requires expensive computations, optical images, and elevation datasets. However, the implementation of orthorectification is easily achievable. The evaluation, hence, would be made—after transferring the ground elevations—by comparing the produced line-of-sight DSM (in terms of misalignment and visual quality) with the achieved integration results through both 2D image registration and orthorectification techniques.

2D image registration, as thoroughly discussed in Goshtasby (2012), consists of four main steps: feature detection, feature matching, mapping function estimation, and image resampling. By having a set of ground control points (GCPs) which are both known in image and object spaces, the need for the first two steps is eliminated. Therefore, the GCPs can be used to estimate the coefficients of the mapping function which are required to find the locations to be resampled in order to register the two datasets. However, there are several mapping functions that can be used to conduct 2D image registration such as affine, projective, Spline, and polynomial transformations. Thus, a few tests, based on check points, need to be made in order to decide which one gives the highest accuracy.

The algorithms of orthorectification perform coordinates transformation between image and ground spaces. This transformation corrects the relief and camera distortions of the input image and converts the image geometry from perspective (or line) to orthographic. Two groups of algorithms are classified based on the transformation direction: direct methods and indirect ones. Direct methods start with a pixel in the image space and calculate its corresponding 3D ground location based on the sensor model information. Conversely, indirect ones start with a ground pixel of the orthoimage and interpolate its elevation from the DEM used. Image resampling to the ground space is an essential step in both methods to build the orthoimage (Novak, 1992; Sheng et al., 2003). Therefore, when the elevation data are available, the satellite orthoimage can be generated directly using the forward RFM model and thus geometrically aligning the optical and the elevation data.

After applying these two selected integration methods, the achieved results are evaluated quantitatively and qualitatively. The quantitative evaluation assesses the misalignment in terms of the Root-Mean-Square-Error (RMSE) of a set of check points (CPs) that are located on the bare-ground. On the other hand, the qualitative evaluation visually assesses the quality of the optical data, i.e., visual image distortions that may have occurred due to resampling.

To avoid any image radiometric and geometric distortions in the employed image-to-image registration technique, the registration must align the elevation data with the optical imagery and not the opposite. Therefore, the qualitative (i.e., visual) assessment is ignored because the original image is preserved. As a result of that, the evaluation should be made quantitatively to assess the horizontal misregistration. Conversely,

orthorectification has a resampling step which needs the RFM model to find the correct image location to be resampled. Thus, the produced orthoimage is required to be assessed qualitatively in terms of its visual quality and quantitatively in terms of RMSE in finding the pixel locations of the CPs.

2.4   Applicability Demonstration

The building detection application was selected to be implemented, in this study, for two reasons: first, to demonstrate the applicability of the developed registration method in a highly demanded remote sensing application, and second, to evaluate the effect of eliminating the misregistration between the optical and elevation data by comparing the detection results, based on the exact same procedure and parameters' values, to a conventional/standard co-registration technique using direct image-to-image registration. Therefore, the following subsections describe the adopted building detection procedure (**Section 2.4.1**), and its evaluation approach (**Section 2.4.2**), to detect the buildings using elevation data integration in an optical VHR dataset of a challenging urban area.

2.4.1   Building Detection Procedure

For the applicability demonstration purposes, the following five steps of the building detection procedure are proposed and justified, with some explanations, to be implemented after achieving the optical-elevation dataset registration:

2.4.1.1 Normalizing the registered elevations

In order to make the off-terrain elevations refer to the same datum, terrain elevations need to be subtracted. Therefore, the normalization could be done in two steps. Firstly,

extracting the terrain elevations from the registered data, and then subtracting these terrain elevations from the original registered data as it is formulated in the following equation:

$$nDSM' = DSM' - DTM' \qquad (2.3)$$

where $nDSM'$ is the normalized line-of-sight elevation data; DSM' is the registered line-of-sight DSM elevations in the image space; DTM' is the extracted terrain elevation data from the registered line-of-sight DSM. The fundamental and most suitable technique for DTM extraction in the image space is to look for local minima elevations within a predefined size of a moving window, then interpolate these identified low elevations to reconstruct a surface which represents approximately the terrain surface (DTM). This ground extraction algorithm is detailed in (Zhang, Tao, & Mercer, 2004).

2.4.1.2   Fusing the VHR image bands

VHR images usually have four multispectral bands (Red, Green, Blue, and Near-Infrared) which have a resolution around one-fourth of that for the panchromatic band; however, these bands provide the color and radiometric information. Therefore, to combine these bands and take the advantages of both types, an image fusion technique needs to be used. The UNB pan-sharpening technique, which is introduced by Zhang (2004), works the best for VHR images because it is designed for new satellite sensors. It fuses the VHR images without distortions and preserves the color information which is critical for the image segmentation technique.

39

2.4.1.3 Segmenting the registered VHR image

To reduce the image complexity and divide it into small objects based on a homogeneity measure of the color information, an image segmentation technique should be used. Furthermore, unlike pixel-based procedures, object-based detection reduces the negative effect of misregistration and facilitates the use of elevation data to support detection (Salehi et al., 2012). As concluded by Dey (2013), multi-resolution segmentation is one of the most appropriate techniques for segmenting VHR images in urban areas. *eCognition®* is the first commercial object-based image analysis software package that implements this algorithm efficiently. Thus, the multi-resolution technique is recommended and adopted for this segmentation step. For more explanation on this technique, readers are encouraged to consult Baatz and Schäpe (2000).

2.4.1.4 Suppressing the scene vegetation

Since the elevations are going to have a crucial role in building detection, other elevated features, such as trees, must be removed to avoid confusion with the building objects. Fortunately, vegetation indices can be used effectively to detect and delineate vegetation. The Normalized Deference Vegetation Index (NDVI), as described in Teillet (1992), is one of the most accurate and popular vegetation indices. The formula to calculate this index is provided in **Equation (2.4)**:

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \tag{2.4}$$

where, *NIR* and *R* are the value of the near infrared and the red bands of the VHR imagery respectively.

2.4.1.5 Detecting the building rooftops

To delineate elevated objects, a thresholding operation based on the normalized elevations needs to be applied. The off-terrain tree objects are already suppressed in the previous step. The best thresholding value can be selected and tested empirically.

2.4.1.6 Finishing the achieved results

To enhance the representation of the detected building objects, post-processing procedures need to be applied. A rule set to enhance the detection results can be developed interactively and effectively using Cognition Network Language (CNL) which is available with the *eCognition®* software. The rule set can be developed on step-by-step bases to achieve the desired results. The set generally comprises steps such as including the segments circumfluent by the detected building objects to the building class, merging the building segments, of short relative border length to non-building segments, to enlarge their areas, and filtering out/excluding the segments of small areas after the merge. Lastly, morphological operations and final editing could be made to enhance the shapes.

2.4.2   Building Detection Evaluation

After implementing the proposed building detection procedure, assessment measures should be calculated. To demonstrate the applicability of the proposed registration method and to evaluate its performance in supporting building detection relative to a standard registration method, the same detection procedure should be implemented again

using a conventional 2D registration method relying only on the accuracy of the registered elevation data.

It is worth emphasising that the concern here in this paper is to measure the relative increase/decrease in the building detection performance, for the tested registration methods, due to only the effect of registration accuracy, not the absolute accuracies for the detection procedure. Therefore, all the steps and their involved parameters should be exactly the same for the conventional registration method being compared with the proposed registration method. Hence, to report the detection performance, assessment measures should be defined.

Completeness, *Correctness* and *Overall Quality* are three widely used measures to assess detection performance. While Completeness (also referred to as Detection rate or as Producer's Accuracy) is the percentage of entities in the reference data that were detected automatically, Correctness (also referred to as User's Accuracy) indicates how well the detected entities match the reference data. However, the Overall Quality of the results provides a compound performance metric that balances Completeness and Correctness. The formulas of these three measures are described by Rutzinger, Rottensteiner, and Pfeifer (2009) as follows:

$$Completeness(Comp.) = \frac{TP}{(TP + FN)} \tag{2.5}$$

$$Correctness\ (Corr.) = \frac{TP}{(TP + FP)} \tag{2.6}$$

$$OverallQuality\ (OQ) = \frac{TP}{(TP + FP + FN)} \tag{2.7}$$

where the true positive (TP) is the number of building objects available in both detection result and the reference data. The false negative (FN) is the number of building objects in the reference dataset that are not detected automatically. The false positive (FP) represents the number of building objects that are detected but do not correspond to the reference dataset. It is worth mentioning that the accuracy assessment can be pixel-based or object-based; however, when the detection is based on segmented objects, pixel-based accuracy measures may be distorted due to problems at the building outlines as found in Rutzinger et al. (2009).

2.5   Data, Experiments and Discussions

2.5.1 Datasets and Study Area

The optical data used in this work are a subset of tri-stereo VHR images acquired by linear sensor with push-broom scanning mode. These overlapped images were taken over the urban area of Melbourne, Australia (Feb. /2012). The two opposite acquisition angles (along-track) of the forward and backward off-nadir images are approximately +15 and -15 degrees respectively. The product ground resolutions are 0.5 m/pixel and 2 m/pixel for the panchromatic and the four multispectral bands respectively. These bands were acquired using the Pleiades-1A satellite sensor. The stereo pairs, along with their relevant sensor model information, are used to photogrammetrically derive elevation data to be employed in the optical-elevation data registration. Two sets of GCPs and CPs, of accuracy within 10-15cm, were used to validate the quality of the stereo-based derived data. **Figurer 2.4** depicts the optical off-nadir VHR image and stereo-based elevation datasets of the study area to be employed in the registration.

The geographic center of the study area, in Melbourne city, is around ($144^{o}57"32'$, -$37^{o}48"55'$). The area contains 170 buildings in about 1.5 Km by 0.8 Km. This test area was selected specifically to represent a dense urban environment with a variety of building shapes, sizes and heights including 15 high-rise ones. About 76% of the scene is buildings (i.e., building density including building facades) and the rest is traffic areas. While the building rooftops represent 46% of the scene, buildings' façades account for 30%.

(a)   (b)
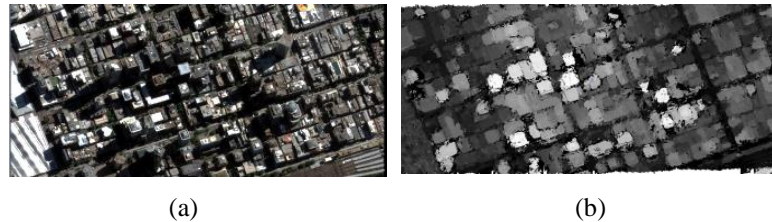
Figure 2.4 The optical and elevation datasets of the study area to be co-registered. (a) Off-nadir VHR satellite imagery, (b) The generated stereo-based elevation data (DSM). The backward off-nadir VHR satellite image was selected to be employed in the optical-elevation data registration and the building detection application.

2.5.2   Registration Results

We compare our proposed method (P-M) with two other optical-elevation data registration methods (explained in **section 2.3.2**): (1) 2D image registration (I-R), and (2) orthorectification (O-R). To execute the 2D image registration method, the relevant mapping function needs to be selected and estimated. Thus, a few tests and comparisons were made to select the most suitable one for the optical and elevation datasets used. Among Affine, polynomials, projective and Spline functions, the 2nd order 2D polynomial produced the best result and hence was selected for the comparison. The results achieved from implementing the I-R, O-R and P-M methods are provided in **Table 2.1** for easy quantitative and qualitative comparison.

Using the I-R method, the original spectral and radiometric image information is preserved because the input imagery was untouched (see **Table 2.1**). However, pixel-by-pixel registration between the two datasets is almost impossible over high-rise buildings in the off-nadir images. The achieved RMSE (h) was of 7.6 pixels based on *ground-level check points*. The effect of such misregistration on a rooftop of a high-rise building would be severe. **Figure 2.5** provides a 3D representation of the achieved optical-elevation data registration using the P-M and I-R methods. It shows the effect of misregistration on top of a high-rise buildings achieved using the I-R method in comparison to the P-M method. Unlike the I-R method, the rooftop is undistorted in the P-M method because of the sub-pixel registration. The original off-nadir VHR image is provided to show the level of the resulted distortion.

The high-rise building roof is preserved due to accurate optical-elevation registration

Distorted high-rise building roof due to optical-elevation misregistration

(a)
The Proposed Method
(P-M)

(b)
The Employed Original VHR Image

(c)
2-D Registration Method
(I-R)
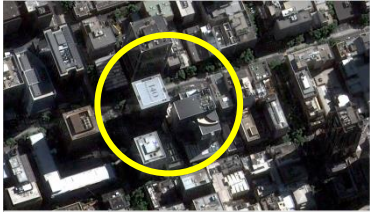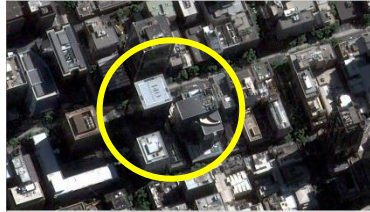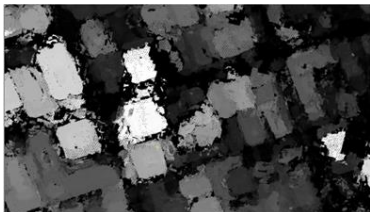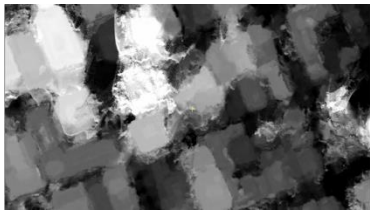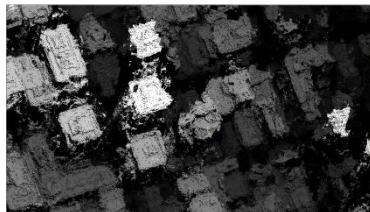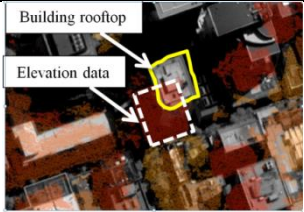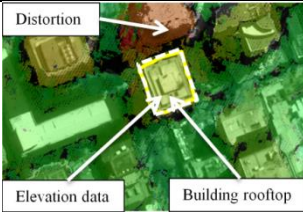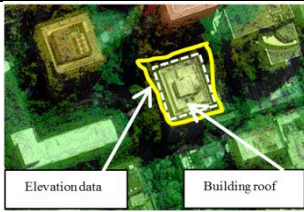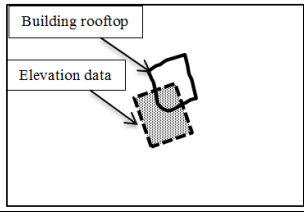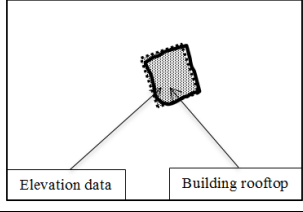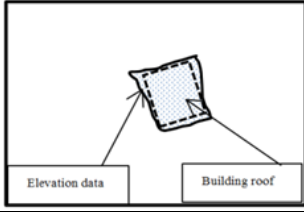
Figure 2.5 3D representation of the registration results achieved for a subset of the datasets used. (a) The registration result achieved by using the proposed method. (b) The original optical off-nadir imagery provided to show the level of distortion. (c) The registration result achieved by using 2nd order 2D polynomial. The yellow color surface represents the terrain (thresholding surface) and hence objects above this surface represent building hypotheses generated in the scene.

On the other hand, O-R registration provides the sub-pixel geometric alignment of RMSE ≤ 0.72 pixel. In comparison to the P-M method, it can be noticed that the O-R and P-M methods both have the same RMSE value for the registration (sub-pixel). This is due to the use of the same sensor model information and the same set of GCPs and CPs. However, the visual quality of the produced orthoimage was distorted, especially at building edges, due to improper resampling and hole-filling (i.e., interpolation) of the imperfect quality of the stereo-based DSM used. When the DSM employed in the orthorectification is independent from the images used (e.g., pre-existing DSM), the occluded areas will produce doubly mapped orthoimages (i.e., so-called ghost images). Hence, a true-orthorectification algorithm must be implemented. However, true-orthoimages usually are not feasible because they are expensive, time consuming, and difficult to achieve.

Among the tested methods, P-M registration, as evidenced in **Table 2.1**, has both advantages of the I-R and O-R methods. In other words, it achieves the accuracy of the O-R method due to the incorporation of the sensor information. Additionally, it keeps the visual quality of the I-R method due to avoiding the original image resampling process. Thus, P-M registration is proven to have sub-pixel registration capability with the preservation of the original image information.

Table 2.1 The quantitative and qualitative results achieved from different optical-elevation data registration methods: image registration (I-R), Ortho-rectification (O-R), and the proposed method (P-M)

Image Registration (I-R)	Ortho-rectification (O-R)	Proposed Method (P-M)
Optical Dataset		
Elevatioin Dataset		

Visuale results: the elevation data are shown on top of the optical data

Color repres.		
B/W repres.		

Achieved Accuracy: RMSE(h)*

7.6 pixels	0.72 pixels	0.72 pixels
Comments - The original image information is preserved. - The misregistration problem is clear over tall off-nadir buildings. - Pixels without elevations are due to the matching failure.	- The original image information is distorted. - No misregistration over tall off-nadir buildings. - Pixels without elevation data were filled by surface interpolation. This makes the orthoimage distortion.	- The original image information is preserved. - No misregistration over tall off-nadir buildings. - Pixels without elevations are due to the matching failure.

* RMSE (h): is the planimetric/ horizontal root-mean-square-error.

2.5.3   Building Detection Results

To demonstrate the applicability of the proposed method, the building detection process was applied, as given in **Section 2.4.1**, on the Melbourne study area. Orthorectification distorts the original radiometric image information, as seen in **Section 2.5.2**, and hence the O-R method was discarded from the comparisons of applying the building detection procedure and evaluating the effect of the optical-elevation data misregistration. This is for two reasons: (a) the segmentations results, in step 3 of **Section 2.4.1**, will be affected directly and negatively by the image distortion particularly over building edges, (b) the misregistration effect is eliminated and hence cannot be evaluated. Therefore, the detection process was executed for the P-M and I-R optical-elevation data registration methods. The detection evaluation was mainly based on the registration accuracy of the elevations in order to evaluate the misregistration effect of the achieved optical-elevation data registration. Therefore, all the implemented steps and their parameters were the same for both the methods applied.

Referring to the proposed detection steps, normalizing the registered elevations ($nDSM'$) was performed after the terrain elevations were extracted. The size parameter of the moving window was selected to be slightly bigger than the maximum building rooftop size in the scene. For the fusion step, no specific values were required. However, three parameters are required to carry out multi-resolution image segmentation. These parameters are known as scale, shape, and compactness and were selected by trial-and-error approach. Concerning the vegetation suppression, this study used an empirically-selected NDVI value of 0.2. Finally, the buildings can be delineated directly by applying a threshold operation.

48

In this study, the buildings were detected by thresholding the normalized heights of the generated line-of-sight DSM. Building objects are polygons (segments) that contain grouped pixels. Each pixel in the building segment has an associated elevation value from the registered line-of-sight DSM which was found to be noisy in many cases. This is related to the quality of the employed elevation data which are extracted by a matching technique (i.e., multi-resolution based normalized-cross-correlation matching). Matching techniques, as it is found in this study, produce mismatches —in the form of sparks (points with extremely high elevations)— in the homogeneous areas like building rooftops, façades, and traffic areas. Therefore, these elevations are averaged for each image segment to reduce this effect. However, it is found that the resulting sparks tend to falsely rise up terrain objects of homogeneous radiometric properties. Consequently, the best threshold value was found empirically equivalent approximately to two storey building height (i.e., 6m).

Figure 2.6 shows a subset of the tested VHR imagery and the reference data (i.e., ground truth) of the building roofs (manually identified) as well as the building detection results achieved by using the I-R registration and by using the P-M method. The results in this figure were achieved by following the same steps and applying the same parameter values for segmentation, normalization, and vegetation suppression. The difference in the detection results is only due to the effect of the registration accuracy between the two methods. The quantitative evaluation of the detection result and the percentage of improvements are provided in **Table 2.2**. To have segments that fit exactly the building edges, most of the buildings in the scene were slightly over-segmented. The percentages

49

in this table were for unrefined results (i.e., no post-processing) for the detected segments that have ≥90% overlap with the reference data.

Table 2.2 The quantitative evaluation of the building detection results achieved by using 2D image registration and the proposed method for optical-elevation registration.

Method	Comp. (%)	Corr. (%)	OQ. (%)
2D image registration (I-R)	84.5	66.1	58.9
Proposed Method (P-M)	89.0	77.5	70.7
Improvement of the P-M	+4.5	+11.4	+11.8

Before discussing these results, it is worth recalling that stereo-based elevations usually suffer from mismatches and matching failure, especially in dense urban areas, which reduce the quality of such models. In this study, the consequence of that appears in the uncompetitive/low detection accuracies achieved since the only information used for the detection is stereo-based elevation. However, although the segmentation result was not optimized and the automatic detection result achieved was not refined, it is clear from **Figure 2.4** that the result of the proposed registration method approaches the reference data and provides a more realistic representation of the detected building rooftops than the I-R method. This was indicated in **Table 2.2** by the difference in the Correctness measure between the two tested methods, which favours the proposed method.

The I-R method provided a Correctness accuracy of about two-third (i.e., 66%), which is relatively low. Regardless of the quality of the stereo-elevation dataset used, since the detection procedure and its parameters were similar between the two methods, the main cause for that is the misregistration effect resulting from the I-R method. This misregistration led to numerous non-building roof objects (e.g., roads between buildings and building façades) being classified falsely as rooftops.

50
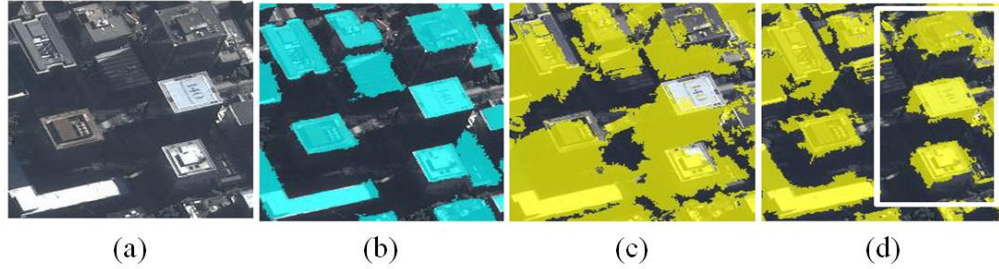
Figure 2.6 A part of the study area contains a few adjacent high-rise buildings. (a) Original off-nadir VHR imagery, (b) reference data of the building roofs, (c) the building detection results achieved using the proposed method, and (d) the building detection results achieved by 2D image registration.

As shown in **Figure 2.6**, the detected rooftops using P-M method were distinguished from each other even when they were very close. In addition to that, the rooftops of the three high-rise buildings were detected correctly despite the large relief displacements. This is due to the accurate registration of the elevation data achieved by developing the line-of-sight DSM which is able to lean with the buildings' lean.

Although the Completeness measure of the I-R method is very close to that of the proposed method, many of the detected objects are false. The factor responsible for that is the shift in the registered elevations, which makes the detection include more false objects. Consequently, the overall quality of the building detection would be affected.

Figure 2.7 illustrates the optical and elevation datasets used along with some intermediate products during the whole implementation process. It shows the registered DSMs by using the 2D image registration technique and the line-of-sight DSM developed by the proposed method side by side. Additionally, it depicts the detected building roofs in the study area, with no post-processing, only based on the registered and normalized elevations.

Optical and Elevation Datasets

Forward VHR off-nadir imagery employed in the optical-elevation data registration

The gennerated stereo-based elevation data (DSM) used in optical-elevation data registration

2D registration Method (I-R)	**Proposed regitration Method (P-M)**

Registered Elevation Data

Registered DSM by conventional 2D registration method

Registered DSM by the proposed method (*line-of-sight DSM*)

Detected buildings (*No post-processing*)

Misregistration effect (*high-rise buildings*)

Facads of the *high-rise buildings* were detected instead of the roofs due to the effect of misregistration

Roofs of *high-rise buildings* were detected correctly due to elimination the effect of misregistration

The refined and edited result of the detected building roofs using the proposed registration method

Figure 2.7 Misregistration effect of the optical-elevation data registration demonstrated through elevation-based building detection using the proposed registration method in comparison with the conventional 2D registration method

It is worth noticing that most of the high-rise building roofs, unlike with P-M method, are not detected when the I-R method was used. This is because the roofs are misaligned with the elevation data as shown in the four magnified high-rise buildings. In contrast, by eliminating the misregistration effect, all of those roofs are detected correctly. Finally, the last part of **Figure 2.7** shows the enhanced and edited buildings detected using the proposed registration method since it yielded the best results. The enhancement was made interactively using CNL-based rule set.

In summary, the detection quality (OQ.) achieved by the proposed method was found to be almost 12 % better than that for the conventional registration method while the other employed parameters in the building detection process were exactly the same. The amount of improvement is mainly attributed to the registration accuracy which assisted the proposed method in detecting the high-rise buildings roofs even with severe relief displacements.

2.6   Conclusions

This paper addressed the problem of elevation data registration with off-nadir VHR optical imagery acquired over a challenging urban area. The paper proposed an accurate method for optical-elevation data co-registration based on developing a line-of-sight DSM solution. This solution is generated by re-projecting the ground elevations back to the image space utilizing the relevant sensor model information. After implementing the whole investigation process, six advantages were identified for the proposed technique:

- It preserves the original image information by performing the resampling in the elevation data.

53

- It provides sub-pixel registration accuracy due to incorporating the relevant sensor model information.

- It is found to be straightforward and inexpensive in terms of elevation data quality (i.e., imperfect DSM with gaps), the computation cost (i.e., no iteration is required for the registration), and the number of the multiple overlapped images required (i.e., can be implemented with even one stereo-pair of VHR images).

- It is applicable to both nadir and off-nadir images regardless of the acquisition angles since it uses the relevant sensor model information.

- It is highly compatible with building detection applications using height information even for off-nadir VHR images acquired over urban areas.

- It increases the usability of off-nadir VHR images since it is highly compatible with elevation-based building detection techniques.

Additionally, the applicability of the proposed method was demonstrated by implementing elevation-based building detection application in the off-nadir VHR image employed to detect buildings with various heights. In comparison to the conventional 2D image registration methods, the proposed technique was found promising in supporting building detection even for high-rise buildings in the off-nadir images where the relief displacements are severe. The improvement achieved in the Overall Quality measure of detecting buildings due to eliminating the misregistration defects, when all other detection steps and parameters are exactly the same, was almost 12 %.

Although the test was made using photogrammetrically derived elevations, the method has a high potential to be successful, with a manipulation step for occlusion,

54

when the elevation model is pre-existing. Therefore, the future work will be to apply this method by employing pre-existing DSMs and test it in a building detection application.

ACKNOWLEDGEMENTS

REFERENCES

Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. And Griesbner,G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*, (pp. 12-23). Wichmann Verlag, Karlsruhe.

Brown, L. (1992). A survey of image registration techniques. *ACM Computing Surveys (CSUR), 24*(4), 325-376. doi: 10.1145/146370.146374

Chen, L., Chiang, T., & Teo, T. (2005). Fusion of LIDAR data and high resolution images for forest canopy modeling. *Proceedings of the 26th Asian Conference on Remote Sensing*, Ha Noi, Vietnam.

Chen, L., Teo, T. , Wen, J. , & Rau, J. (2007). Occlusion compensated true orthorectification for high-resolution satellite images. *Photogrammetric Record, 22*(117), 39-52.

Chen, L., Zhao, S., Han, W., & Li, Y. (2012). Building detection in an urban area using lidar data and QuickBird imagery. *International Journal of Remote Sensing, 33*(16), 5135-5148.

Dey, V. (2013). Image segmentation techniques for urban land cover segmentation of vhr imagery: recent developments and future prospects. *International Journal of Geoinformatics, 9*(4), 15-35.

Dowman, IJ. (2004). Integration of LIDAR and IFSAR for mapping. *International Archives of Photogrammetry and Remote Sensing, 35*(B2), 90-100.

Fonseca, L. , & Manjunath, B. (1996). Registration techniques for multisensor remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing, 62*(9), 1049-1056.

Fraser, C., & Hanley, H. (2003). Bias compensation in rational functions for IKONOS satellite imagery. *Photogrammetric Engineering and Remote Sensing, 69*(1), 53-58.

Fujii, K., & Arikawa, T. (2002). Urban object reconstruction using airborne laser elevation image and aerial image. *IEEE Transactions on Geoscience and Remote Sensing, 40*(10), 2234-2240.

Goshtasby, A. A. (2012). *Image registration: Principles, tools and methods*. Springer Science and Business Media. London.

Grodecki, J., & Dial, G. (2003). Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering and Remote Sensing, 69*(1), 59-68.

Habib, A., & Al-Ruzouq, R. (2005). Semi-automatic registration of multi-source satellite imagery with varying geometric resolutions. *Photogrammetric Engineering and Remote Sensing, 71*(3), 325-332.

Habib, A., Kim, E., & Kim, C. (2007). New methodologies for true orthophoto generation. *Photogrammetric Engineering and Remote Sensing, 73*(1), 25-36.

Habib, A., Morgan, M., Jeong, S., & Kim, K. (2005). Analysis of epipolar geometry in linear array scanner scenes. *The Photogrammetric Record, 20*(109), 27-47. doi: 10.1111/j.1477-9730.2005.00303.x

Heinl, M., Walde, J., Tappeiner, G., & Tappeiner, U. (2009). Classifiers vs. input variables—the drivers in image classification for land cover mapping. *International Journal of Applied Earth Observation and Geoinformation, 11*(6), 423-430.

Hong, G., & Zhang, Y. (2008). Wavelet-based image registration technique for high-resolution remote sensing images. *Computers and Geosciences, 34*(12), 1708-1720.

Karras, G., Grammatikopoulos, L., Kalisperakis, I., & Petsa, E. (2007). Generation of orthoimages and perspective views with automatic visibility checking and texture blending. *Photogrammetric Engineering and Remote Sensing, 73*(4), 403-411.

Khosravi, I., Momeni, M., & Rahnemoonfar, M. (2014). Performance evaluation of object-based and pixel-based building detection algorithms from very high spatial resolution imagery. *Photogrammetric Engineering and Remote Sensing, 80*(5), 519-528.

Lehner, M, & Gill, R. (1992). Semi-automatic derivation of digital elevation models from stereoscopic 3-line scanner data. *European space agency-publications-esa sp, 347*, 263-263.

Mastin, A., Kepner, J., & Fisher, J. (2009). Automatic registration of LIDAR and optical images of urban scenes. *Proceedings of the IEEE Conference on the Computer Vision and Pattern Recognition(CVPR 09)*, Miami, FL, USA.

Mishra, R., & Zhang, Y. (2012). A review of optical imagery and airborne lidar data registration methods. *The Open Remote Sensing Journal, 5*, 54-63.

Novak, K. (1992). Rectification of digital imagery. *Photogrammetric Engineering and Remote Sensing, 58*(3), 339-344.

Palenichka, R., & Zaremba, M. (2010). Automatic extraction of control points for the registration of optical satellite and LiDAR images. *Geoscience and Remote Sensing, IEEE Transactions on, 48*(7), 2864-2879.

Rutzinger, M., Rottensteiner, F., & Pfeifer, N. (2009). A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2*(1), 11-20.

Salehi, B., Zhang, Y., & Zhong, M. (2011). Object-based land cover classification of urban areas using VHR imagery and photogrammetrically-derived DSM. *Proceedings of the ASPRS Annual Conference*, Milwaukee, WI, USA.

Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). A review of the effectivness of the spatial information used in urban land cover classification of vhr imagery. *International Journal of Geoinformation, 8*(2), 35-51.

Sefercik, U., Karakis, S., Bayik, C., Alkan, M., & Yastikli, N. (2014). Contribution of normalized dsm to automatic building extraction from hr mono optical satellite imagery. *European Journal of Remote Sensing, 47*, 575-591.

Sheng, Y., Gong, P., & Biging, G. (2003). True orthoimage production for forested areas from large-scale aerial photographs. *Photogrammetric Engineering and Remote Sensing, 69*(3), 259-266.

Stankov, K., & He, D. (2014). Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(10), 4069-4080. doi: 10.1109/JSTARS.2014.2308301.

Suliman, A., & Zhang, Y. (2014). Integration of off-nadir VHR imagery with elevation data for advanced information extraction. *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Quebec, Canada.

Tao, C., & Hu, Y. (2001). A comprehensive study of the rational function model for photogrammetric processing. *Photogrammetric Engineering and Remote Sensing, 67*(12), 1347-1358.

Teillet, P. (1992). An algorithm for the radiometric and atmospheric correction of AVHRR data in the solar reflective channels. *Remote Sensing of Environment, 41*(2), 185-195.

Weidner, U., & Förstner, W. (1995). Towards automatic building extraction from high-resolution digital elevation models. *ISPRS journal of Photogrammetry and Remote Sensing, 50*(4), 38-49.

Xiong, Z., & Zhang, Y. (2010). A critical review of image registration methods. *International Journal of Image and Data Fusion, 1*(2), 137-158.

Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering and Remote Sensing, 70*(6), 657-661.

Zhang, Y., Tao, C., & Mercer, J. (2004). An initial study on automatic reconstruction of ground DEMs from airborne IfSAR DSMs. *Photogrammetric Engineering and Remote Sensing, 70*(4), 427-438.

Zitová, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing, 21*(11), 977-1000. doi: 10.1016/S0262-8856(03)00137-9

Chapter 3: REGISTRATION-BASED MAPPING OF ABOVEGROUND DISPARITIES (RMAD) FOR BUILDING DETECTION IN OFF-NADIR VHR STEREO SATELLITE IMAGERY[1]

Abstract

Reliable building delineation in very high resolution (VHR) satellite imagery can be achieved by precise disparity information extracted from stereo pairs. However, off-nadir VHR images over urban areas contain many occlusions due to building leaning that creates gaps in the extracted disparity maps. The typical approach to fill these gaps is by interpolation. However, it inevitably degrades the quality of the disparity map and reduces the accuracy of building detection. Thus, this research proposes a registration-based technique for mapping the disparity of off-terrain objects to avoid the need for disparity interpolation and normalization. The generated disparity by the proposed technique is then used to support building detection in off-nadir VHR satellite images. Experiments in a high-rise building area confirmed that 75% of the detected building roofs overlap precisely the reference data, with almost 100% correct detection. These accuracies are substantially higher than those achieved by other published research.

3.1   Introduction

3.1.1 Overview and Motivation

Building detection is important for many applications. City planning and management (Nielsen & Ahlqvist, 2014), city modeling (Singh, Jain, & Mandla, 2014), population estimation (Xie, Weng, & Weng, 2015; Zhu, Li, Liu, & Fu, 2015), and urban growth monitoring (Sugg, Finke, Goodrich, Susan Moran, & Yool, 2014) are just a few examples where continuously updated building information is required. The very high resolution (VHR) satellite images with their large coverage and detailed spatial information provide substantial information necessary for mapping complex urban environments. Therefore, building detection and mapping using VHR satellite images have become an active research area in the remote sensing community.

VHR satellite sensors usually acquire off-nadir images with across-track and/or along-track angles using various acquisition modes (e.g., target, corridor, and stereo acquisition modes). The readily available off-nadir satellite images captured in a stereo mode allow generating elevation data to support building detection at a lower cost than other sources of elevation data such as LiDAR (Light-detection-And-Ranging). However, off-nadir images always suffer from the compounded effect of tilt and relief displacement for all elevated objects. This effect reveals building façades, hides areas, and creates registration difficulty between elevation surface models and the images.

Building detection methods based on optical images make use of various types of information provided by VHR images. These information types can be categorized broadly into two classes: mono-based and stereo-based image information. The mono-

based image information includes spectral and spatial (morphological, textural, and contextual) information that is extracted directly from VHR images. In contrast, the stereo-based information comprises image-derived products such as digital surface models (DSMs) that are generated after photogrammetric processing of multiple overlapped images (Salehi, Zhang, Zhong, & Dey, 2012). Following this taxonomy, building detection methods can be divided into two categories: mono-based and stereo-based.

The most successful building detection methods from VHR images using mono-based image information are comprehensively evaluated by Khosravi, Momeni, and Rahnemoonfar (2014). However, mono-based image information poses some limitations for building detection. Chen, Shang, and Wu (2014) reported a few challenges for building detection methods and stated that the intra-similarity of most impervious surfaces in VHR images is the most intractable difficulty for such methods. For instance, it is very difficult to automatically distinguish building roofs from parking lots with the same spectral and spatial (i.e., textural and morphological) properties. To overcome this deficiency, Ghaffarian and Ghaffarian (2014) introduced a shadow-based building detection method (i.e., a method based on a type of contextual information) in VHR images. Although the results are promising, they concluded that such methods are still incapable of separating spectrally-spatially similar objects. Therefore, when employing off-nadir VHR images over dense urban areas, the challenges of mono-based building detection methods will increase due to the existence of façades and the possibility of buildings overshadowing others.

Considering that buildings are elevated impervious surfaces, stereo-based information provides the third dimension as the key feature for accurate detection and reliable differentiation from similar traffic areas. Heinl, Walde, Tappeiner, and Tappeiner (2009) found that the accuracy of urban mapping is increased when the elevation data are incorporated regardless of the technique implemented. Krauß, D'Angelo, Kuschk, Tian, and Partovi (2015) and Tian, Cui, and Reinartz (2014) are just two successful examples of stereo-based building detection approaches. The third dimension in these examples was the surface elevation as derived photogrammetrically from stereo images.

The incorporation of elevation data in building detection requires a co-registration between the VHR optical images and the corresponding DSMs. However, direct integration of these two datasets often introduces a misregistration problem (Van de Voorde, De Genst, & Canters, 2007). For highly elevated buildings in off-nadir images, the misregistration is a serious and challenging problem due to the severe relief displacement of such buildings. This particular problem is thoroughly discussed in Suliman and Zhang (2015).

To circumvent this problem, one may use disparity information as the alternative stereo-based third dimension. Disparity maps are constructed by executing an image matching technique to measure the distance, in pixels, between each pixel in one image and its conjugate in the other stereo mate along the epipolar direction. The advantage of using disparity maps is that they, by definition, have exactly the same reference frame as one of the stereo images. Replacing elevation models with disparity maps avoids several computationally expensive steps (e.g., aerial triangulation and accurate co-registration)

required for implementing building detection methods using stereo-based elevations. The following section is a brief review of building detection methods based on disparity information proposed in the existing literature.

3.1.2 Related Works and Challenges

Disparity-based building detection methods rarely appear in research publications. Oriot (2003) proposed a technique based on segmenting the disparity map into building and background classes. Beumier (2008) introduced a simple and fast technique to detect buildings in directly acquired epipolar Ikonos stereo pairs. The technique uses the disparity of building edges for the detection. Unlike these two techniques that assume flat terrain, an approach proposed by Krauß and Reinartz (2010) for urban object modeling was based on fusing disparity maps with VHR optical stereo data. The approach includes an appealing technique to extract terrain disparity. More recently, Krauß, Sirmacek, Arefi, and Reinartz (2012) made use of the classification power of eight-band VHR stereo images from the WorldView-2 sensor to complete and enhance the generated disparity map. Although these reviewed methods are promising, two major challenges are identified: occlusion effect and terrain variation effect. These two difficulties become extremely challenging when dealing with off-nadir VHR images captured over dense and non-flat urban areas.

Occlusion is the result of off-nadir acquisition angles which create the leaning appearance of buildings in VHR images. It is impossible to find point matches in these hidden areas and, therefore, to measure the disparity. The consequence is disparity maps with many no-data regions that are normally filled by interpolating the surrounding data.

However, interpolation in urban areas results in over smoothing building boundaries and missing narrow roads. This misleading information destroys the quality of the generated disparity map and affects the subsequent processes. To minimize the occlusion effect, the reviewed literature dictates that the convergence angle of the stereo images should be small to guarantee high similarity of the overlapped images. Another possible solution is to use multiple stereo pairs to eliminate the occlusion effect. Unfortunately, these two options are not always available.

The other challenge that faces disparity-based building detection methods is the need to remove bare-earth effects. Terrain elevations may cause buildings with the same aboveground height to have different disparity values. Thus, a terrain disparity map (TDM), representing the bare earth, needs to be extracted and subtracted from its corresponding surface disparity map (SDM) that describes the visible surface. The result is a normalized surface disparity map (nSDM) that represents only the objects above the extracted TDM (i.e., $nSDM = SDM - TDM$).

To identify bare earth, an algorithm that uses different ranks of percentile filters to approximate the terrain variation is introduced by Weidner and Förstner (1995). This empirical algorithm assumes that the ground-level areas dominate the scene. Arefi, Engels, Hahn, and Mayer (2007) developed the geodesic algorithm that iteratively executes a dilation process until a predefined marker surface is met. However, this algorithm requires repetitive computation and the results vary depending on the selected surface. Krauß and Reinartz (2010) proposed a steep-edge algorithm that uses the subtraction of two sizes of median-filter results to detect the areas at the bottom of steep

walls. It assumes that the detected areas are on the ground and not occluded. However, in dense urban areas, different levels of building roofs may be adjacent to each other in a way that satisfies the condition of the steep-edge algorithm, thus produces false surface information.

Being based on specific assumptions, these algorithms have limitations in extracting terrain from incomplete urban disparity maps especially when the occlusion in off-nadir images is serious. In contrast, the local-minima technique developed by Zhang, Tao, and Mercer (2004) is a more general technique without prior assumptions. It is based on interpolating ground-level points which are detected by a moving window of a constant size that looks for local minima values.

In summary, off-nadir VHR stereo images are burdened by inherent occlusion at building edges when acquired over dense urban areas. If the convergence angle of the stereo images is not small enough, the implemented interpolation technique to fill the gaps will result in misleading terrain disparities over narrow streets and between adjacent buildings that destroy the quality of the subsequent process.

3.1.3   Research Objectives and Hypothesis

The ultimate aim for this research is detecting building roofs in off-nadir VHR satellite images acquired over a dense and reasonably non-flat urban area. The adopted approach is based on using normalized disparity data derived from a stereo pair. Thus, crucial to this study is developing a technique for generating normalized disparity maps. For that purpose, we propose that if the original stereo images are (1) rectified to eliminate the y-direction disparity (y-parallax) of all corresponding pixels (thereby

creating epipolar images) and (2) co-registered with the corresponding ground-level objects (e.g., roads) to eliminate the x-disparity of the terrain, then the remaining measurable x-disparity should represent only the off-terrain objects (i.e., nSDM). Consequently, both the interpolation step (to fill data gaps caused by occlusion) and the terrain extraction process (to normalize the SDM) will be bypassed.

The novelty of this technique is in the concept of co-registering terrain-level objects to directly measure aboveground disparity information without the need for applying either interpolation or data normalization. An earlier version of this technique was presented in Suliman, Zhang, and Al-Tahir (2016). The developed technique is then incorporated in the actual task of stereo-based building detection using disparity data. The details of the proposed technique and its implementation are described in **Section 3.2** of this paper.

Section 3.3 describes the validation procedure to evaluate the performance of the proposed technique in building detection relative to similar results based on published methods for epipolar rectification, gap surface interpolation, and stereo data normalization. **Section 3.4** provides all of the experimental results achieved in this study. **Section 3.5** compares and discusses the results, while **Section 3.6** draws the conclusions.

3.2   Methodology

The methodology for this research aims at tackling the two tasks stated before: generating a normalized disparity map and incorporating it in the detection of buildings. These two tasks are followed by a validation procedure, as detailed in **Section 3.3**, for the performance of detecting buildings based on the generated disparity.

3.2.1 Registration-based Technique for Mapping Aboveground Disparities (RMAD)

The key concept for the proposed technique is to process the original VHR stereo images in a way that allows image matching techniques to measure the disparity of aboveground objects only. This process includes four steps: (1) minimizing the relief distortion caused by the terrain in the stereo images, (2) removing the y-disparity between the stereo images, (3) co-registering terrain-level features and, hence, eliminating their x-disparity, and (4) mapping the disparity of the remaining x-disparity values that correspond to off-terrain objects. These steps are further discussed and justified in the following subsections.

3.2.1.1 Minimizing terrain relief displacement

The variation in the terrain elevation produces relief distortions in any acquired image. Because of the different viewing angles for a stereo pair of satellite images, these distortions make the relative distance between any two points vary from one image to another. **Figure 3.1** describes this effect by showing the difference in the distance between two points in the two image planes. It is clear that a straightforward co-registration of the terrain-level features in a stereo pair is impossible. Typically, relief distortion can be completely corrected for if a terrain model of the imaged area is available. However, since such models are not always available, minimization of this distortion should be considered instead.

67

Figure 3.1 Label text within figure:

Linear Sensor

Image plane a b a' b'

ab

$a'b'$

ab

Terrain relief displacement

Light rays

A

Average elevation

B

Projection plane

la lb

Terrain relief displacement ($la - lb$) is minimum
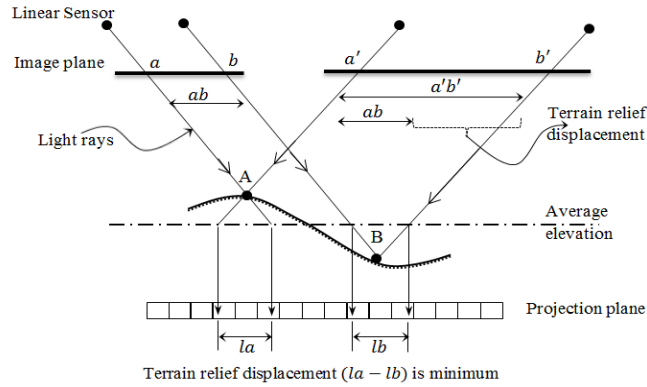
Figure 3.1 Minimization of terrain relief displacement

Dense urban areas are characterized by reasonable road grades based on engineering design standards. For instance, the standards of the American Association of State Highway Transportation Officials (AASHTO) limit the maximum grades of urban areas at 50 km/h design speed to the range of 7 to 12% depending on the terrain and road class (AASHTO, 2001). The resulting relief effects from such road grades can be corrected by re-projecting the images onto a common object-space plane that minimizes the relief displacements. Such a plane is defined by the average elevation of the terrain variation, as demonstrated graphically in **Figure 3.1**.

The image re-projection process can be executed by (1) finding the approximate ground coverage of the stereo images from their meta-data, (2) constructing a 2D grid with a spacing distance usually similar to the ground resolution of the images, and (3) projecting the stereo images to the plane of the average terrain elevation using the relevant sensor model information. We strongly recommend that the resampling process of this step is to be done in the image space for better visual quality of the output images. This is achievable through the use of ground-to-image (forward) rational function model

(RFM), which is the most popular generalized sensor model for VHR satellite images (Tao & Hu, 2001).

3.2.1.2 Eliminating Y-disparity of the stereo-pair

Eliminating the y-disparity between the stereo images makes all conjugate points in the left image lie on the same line in the right image. This situation is realized by rectifying the stereo images into the epipolar geometry. Linear array sensors used in all VHR satellite systems acquire scenes with a different perspective center for each image column. This effect destroys the straightness of the epipolar lines as comprehensively discussed by Habib, Morgan, Jeong, and Kim (2005). However, it is concluded by Wang, Hu, and Li (2011) that the epipolar curves of satellite stereo images can be approximated by straight lines if the images are projected onto a common projection reference plane (PRP) in the object space (PRP epipolarity model). Thus, to maintain the terrain relief distortion at minimum, the common projection plane in the object space must be at the average terrain elevation of the imaged area. Consequently, our study adopts the average terrain level for the PRP epipolarity model to eliminate the y-disparity of the stereo-pair and to minimize the terrain relief distortions.

Following this step of image projection, the stereo images need to be relatively reoriented along the epipolar direction to make the corresponding points lie on the same image row. This epipolar direction should be identified before applying the relative orientation. A practical approach to calculate the epipolar direction is to match a set of distinct points on the tops of high buildings in one projected image to their conjugates in the other image. The object-space PRP locations of any of these points can be used

directly to calculate the epipolar direction in the object-space PRP plane as in **Equation 3.1**.

$$\theta = \tan^{-1}\left(\frac{Y_2 - Y_1}{X_2 - X_1}\right) \tag{3.1}$$

where (X_1, Y_1) and (X_2, Y_2) are the object-space PRP coordinates of the point in the two projected stereo images. A rotation by the amount of angle θ can then be applied to the two projected images around their centers to make their rows parallel to the epipolar direction.

A vertical shift transformation should also be applied to align the coordinate systems for both images. This shift can be calculated using the coordinates of one point pair in both of the oriented images. The whole process is described in **Figure 3.2** which illustrates the applied rotation around image centers. The product of this step is an epipolar stereo pair with approximated straight epipolar lines, and all conjugate points lie on the same row in both images.
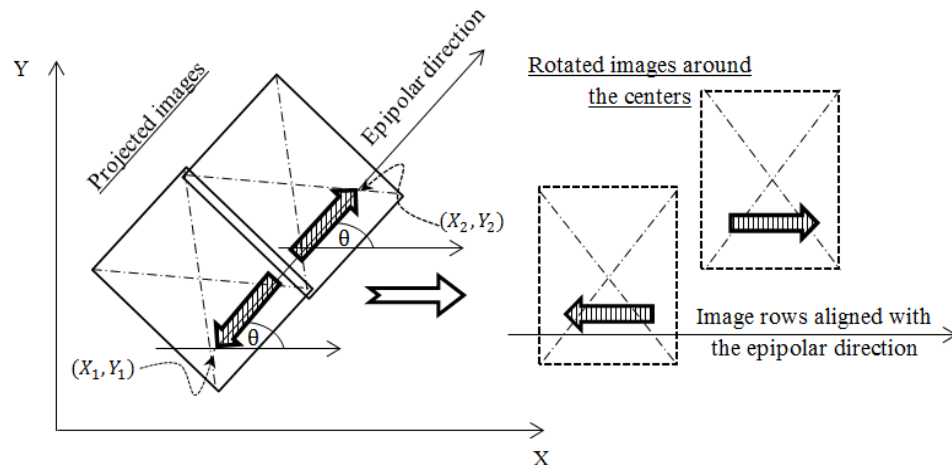
Figure 3.2 image reorientation to make the image rows parallel to the epipolar direction.

3.2.1.3   Co-registering of terrain-level features

Following the argument of this research, x-disparity measured in epipolar resampled stereo images would only represent the off-terrain objects if corresponding ground-level objects are co-registered. Accordingly, a shift is required to geometrically align the terrain-level objects (e.g., roads) in both images, and thus, eliminate their x-disparity. This alignment must maintain the epipolar condition already achieved in the previous step. Therefore, the alignment must be in the row direction only.

The required transformation for this step is a shift in the x-direction (epipolar direction/rows) of the right epipolar image, for example, with respect to the left one. Therefore, by having an accurately matched ground-level point pair, the x-shift can be calculated and applied to one epipolar image (e.g., the right image). To guarantee the matched point in this step to be of ground level, manual matching is recommended due to its feasibility. Once the point is identified, determining the x-shift is straightforward. This shift can be implemented simultaneously with the y-shift mentioned in the previous step as described in **Equation 3.2**:

$$I(x', y') = I(x, y) + S(\Delta x, \Delta y) \tag{3.2}$$

where $I(x', y')$ is one of the projected images reoriented along the epipolar direction, $I(x, y)$ is the same image after co-registration using the shifts in the x-direction(Δx) and y-direction(Δy) calculated from a point pair on the ground level.

3.2.1.4 Mapping the disparity of the aboveground objects

The whole process of the proposed RMAD technique is flowcharted in **Figure 3.3**. The last step in RMAD is the generation of a disparity map for the aboveground features by applying an image matching technique based on epipolar constraint.

There are several alternatives for an image matching algorithm to be implemented for creating disparity maps. Alobeid, Jacobsen, and Heipke (2010) provided a critical review of the state-of-the-art matching algorithms for dense disparity map generation. They concluded that the most suitable matching algorithm for urban areas is the semi-global matching (SGM) technique as introduced in Hirschmüller (2008) due to several advantages. For instance, it provides a sub-pixel matching accuracy and preserves the discontinuity at the building edges. Furthermore, it is insensitive to the illumination and reflection changes. Consequently, this study adopts this algorithm for the generation of the disparity information.
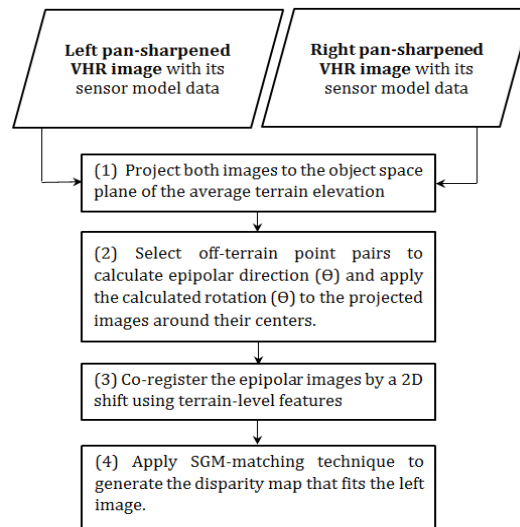
Figure 3.3 The work flow of disparity maps generation for the off-terrain objects (RMAD technique).

3.2.2 Disparity-based Building Detection

A procedure can now be designed for building detection using the pan-sharpened VHR image along with the accurately co-registered disparity map. The multi-spectral bands of the VHR image are used to define the borders of the scene features, while the normalized disparity information identifies the aboveground objects (e.g., buildings and trees). Based on that, the general steps for a disparity-based building detection procedure can be proposed as follows:

3.2.2.1 Image segmentation

Image pixels are grouped into classes based on the homogeneity measure of color information. The resulting segments represent meaningful objects in the scene. Several segmentation techniques are introduced in the literature. However, multi-resolution segmentation (Baatz & Schäpe, 2000) is one of the most appropriate techniques for segmenting VHR images in urban areas as concluded by Dey (2013). For improved segmentation results, the VHR image is required to be pan-sharpened. The UNB pan-sharpening technique, introduced by Zhang (2004), works best for VHR images of the new satellite sensors because it preserves color information that is critical for all radiometric-based segmentation techniques.

3.2.2.2 Vegetation suppression

The disparity values have a crucial role in the subsequent steps for building detection. Thus, other elevated features, such as trees, must first be removed to avoid confusion with buildings. Vegetation indices based on red and infra-red bands of VHR

images are used effectively to detect and delineate vegetation. One of the most accurate and popular vegetation indices is the Normalized Difference Vegetation Index (NDVI).

All image segments related to trees and grass objects can be identified by applying a thresholding operation to the calculated NDVI values as in **Equation 3.3**, and subsequently, their corresponding disparity values are suppressed (set to zero).

$$\text{Disp}(x,y) = \begin{cases} 0 & if\ NDVI\ \geq t \\ Disp(x,y) & Otherwise \end{cases} \tag{3.3}$$

where $\text{Disp}(x,y)$ is the disparity value at the x and y location in the object space plane; t is the NDVI threshold value.

3.2.2.3 Disparity thresholding

At this stage, building objects can be detected directly and without confusion by simply applying a threshold to the disparity map. This is due to the fact that, after the last step, the remaining disparity values will only represent buildings. Furthermore, in an off-nadir VHR stereo pair, the disparities should represent buildings' rooftops only without any buildings' façades. This takes advantage of the geometry of image acquisition that makes building façades in one stereo image not appear in the other one (**Figure 3.4**). Because of the façades' dissimilarity and/or the imposed pre-specified search range in the SGM algorithm, pixels that belong to building façades cannot normally be matched in an off-nadir stereo pair.

74
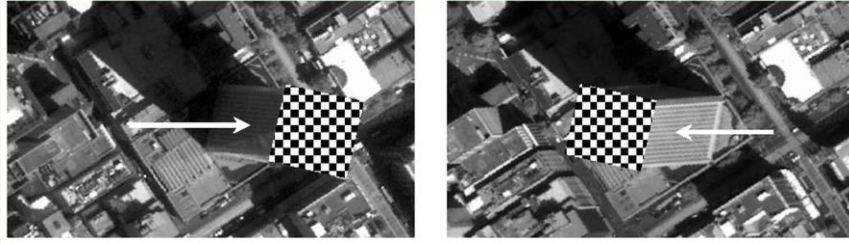
Figure 3.4 Building façades in an off-nadir VHR stereo pair. Due to the different viewing angles for VHR linear sensors, building façades that appear in the backward image (left) do not appear in the forward one (right).

3.2.2.4 Result finishing

A post-processing procedure may be applied to enhance the representation of the detected buildings. The procedure includes removing the noise, filling the holes, and merging the detected segments. Object-based image analysis software packages (e.g., *eCognition*®) provide user friendly and efficient tools for such finishing purposes.

The whole proposed approach for disparity-based building detection described above is flowcharted in **Figure 3.5**. The bolded parallelogram in this figure underscores that the critical input for detecting buildings is the aboveground disparity map.
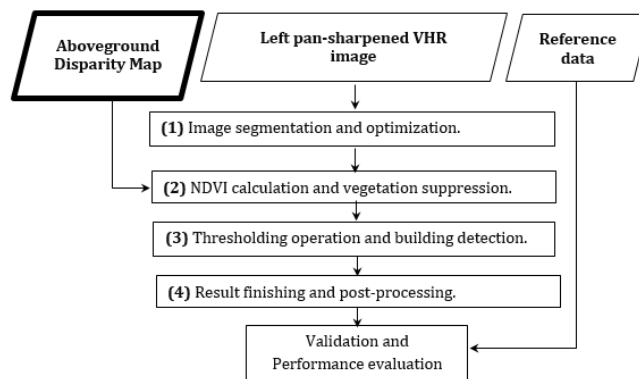
Figure 3.5 The work flow for stereo-based building detection using normalized disparity information.

75

3.3   Validation Procedure

The main intent of this section is to prove the efficiency of the developed procedure as well as its advantages. Accordingly, two tests are devised to validate the results of the technique developed in this research. The first is to evaluate the building detection performance relative to reference data; the second is to evaluate the results achieved by the developed RMAD technique against those achieved by closely related published techniques. The proposed evaluations are detailed in the following subsections.

3.3.1   Accuracy Assessment of Building Detection

Accuracy is assessed by evaluating the detected buildings' roofs against a reference dataset after implementing the developed building detection procedure. *Completeness*, *Correctness*, and *Overall Quality* are three widely used measures to assess detection performance. These detection measures are an adapted version of the classification accuracy assessment measures originally introduced by Story and Congalton (1986) which are referred to as Producer's accuracy, User's Accuracy, and Overall accuracy respectively. Completeness is the percentage of the entities in the reference data that are correctly detected. Correctness indicates how well the detected entities match the reference data. The Overall Quality of the results provides a compound performance metric that balances Completeness and Correctness. The formulas of these three measures are described by Rutzinger, Rottensteiner, and Pfeifer (2009) as follows:

$$Completeness(Comp.) = \frac{TP}{(TP + FN)} \tag{3.4}$$

$$Correctness\ (Corr.) = \frac{TP}{(TP + FP)} \tag{3.5}$$

$$OverallQuality\ (OQ) = \frac{TP}{(TP + FP + FN)} \tag{3.6}$$

where true positive (TP) is the number of correctly identified building roof segments. The false negative (FN) is the number of building roof segments in the reference dataset that are not detected. The false positive (FP) represents the number of building roof segments that are detected but do not correspond to the reference dataset. It is worth mentioning that the detection assessment can be pixel-based or object-based (Rottensteiner, Trinder, Clode, & Kubik, 2005; Rutzinger et al., 2009). However, the definition for these measured entities in this study represents the total number of pixels that are labeled as roof (i.e., the total area in pixel units).

3.3.2   Comparative Evaluation of Building Detection

Following the main argument for this study, the procedure for building detection centered around RMAD-generated disparity map bypasses the normally required interpolation step (to fill data gaps caused by occlusion) and the terrain extraction process (to normalize the SDM). As a result, there is a need to evaluate the impact of circumventing these two steps on the quality of building detection. Accordingly, building detection results based on the developed RMAD technique are compared to the results obtained from a disparity map generated based on an RFM-based epipolar image pair. This RFM epipolarity model, as introduced by Zhao, Yuan, and Liu (2008), creates epipolar rectified images without considering the changes in terrain relief distortions.

The RFM-based epipolarity technique uses a pre-selected set of evenly distributed object space elevations and the two forms of RFM model: the image-to-ground (backward) model and the ground-to-image (forward) model. For an image pixel in the left image, the former model is used to calculate the corresponding planimetric object-space coordinates for a set of elevations. Then, the latter model is used to project the calculated 3D object-space coordinates to the image-space of the right image. By repeating this process for a set of evenly distributed elevations, the epipolar images can be resampled without rectifying the terrain relief distortions (Zhao et al., 2008).

Following the creation of the epipolar pairs, the next step is to generate the corresponding surface disparity map (SDM) using the SGM algorithm. An interpolation step is required afterward to fill the gaps in the disparity map caused by the occlusion effects. This is followed by the extraction of the terrain level disparity model (TDM) since terrain relief distortions are not rectified by the RFM-based epipolarity technique. The terrain extraction technique selected for this step is the local minima technique developed by Zhang et al. (2004) because it is the most successful one among the reviewed techniques.

Subsequently, the TDM is subtracted from the SDM to obtain the normalized disparity map (nSDM) which is then used in the disparity-based building detection procedure described in **Section 3.2.2**. Finally, the detected buildings' rooftops are evaluated against the reference data. **Figure 3.6** depicts the work flow for the steps described in the previous paragraphs and compares them against the developed RMAD technique for the disparity-based building detection (in dashed lines).

78
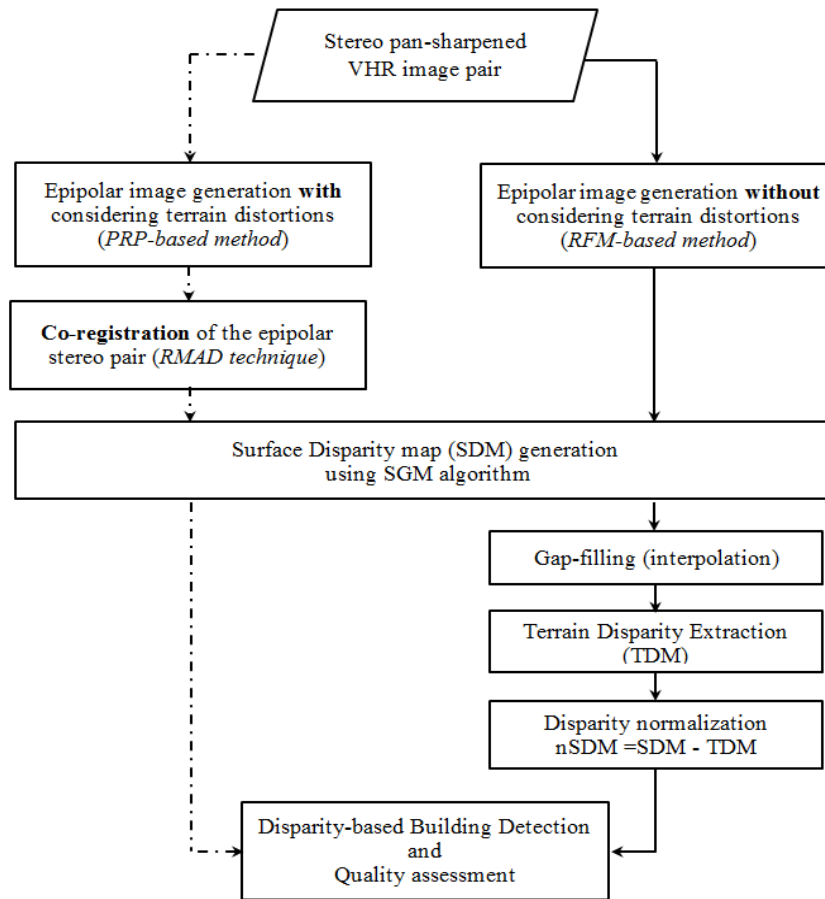
Figure 3.6 The work flow for building detection using RMAD-based (dashed line) versus RFM-based (continuous line) disparity map.

3.4   Experimental Results

3.4.1   Dataset and Study Area

The optical data used in this work are a subset of stereo pair of VHR images acquired by linear sensor with push-broom scanning mode. These stereo images were taken by Pleiades-1A satellite over the urban area of Melbourne, Australia (2012). The two opposite acquisition angles of the forward and backward off-nadir images are

approximately +15 and -15 degrees respectively. The product's ground resolutions are 0.5 m/pixel for the panchromatic and 2 m/pixel for the four multispectral bands.

This test area was selected specifically to represent a dense urban environment with a variety of building shapes, sizes, and heights. The study area spans about 1.5Km by 0.8Km and contains 170 buildings, 15 of which are high-rise buildings. About 76% of the scene is pixels belonging to buildings (46% of the scene is rooftops and 30% is building façades) while the rest represent vegetation and traffic areas. A reference dataset for building rooftops in the scene was manually produced based on expert's interpretation of the satellite image.

3.4.2   Results of the Developed RMAD Technique

The steps in the developed RMAD technique were implemented and applied as detailed in **Section 3.2** on the study area. The derived results, including co-registration, disparity map generation, and the detected buildings are evaluated in the following sections.

3.4.2.1   Results of co-registration and disparity map generation

In this step, the right image of the epipolar pair was co-registered to the left one by applying a 2D translation calculated from a set of ground-level points. In our implementation, four points defined by intersection of road centers were identified and used. The details of these points and the result of co-registering the stereo epipolar pair are all illustrated in **Figure 3.7**.

80
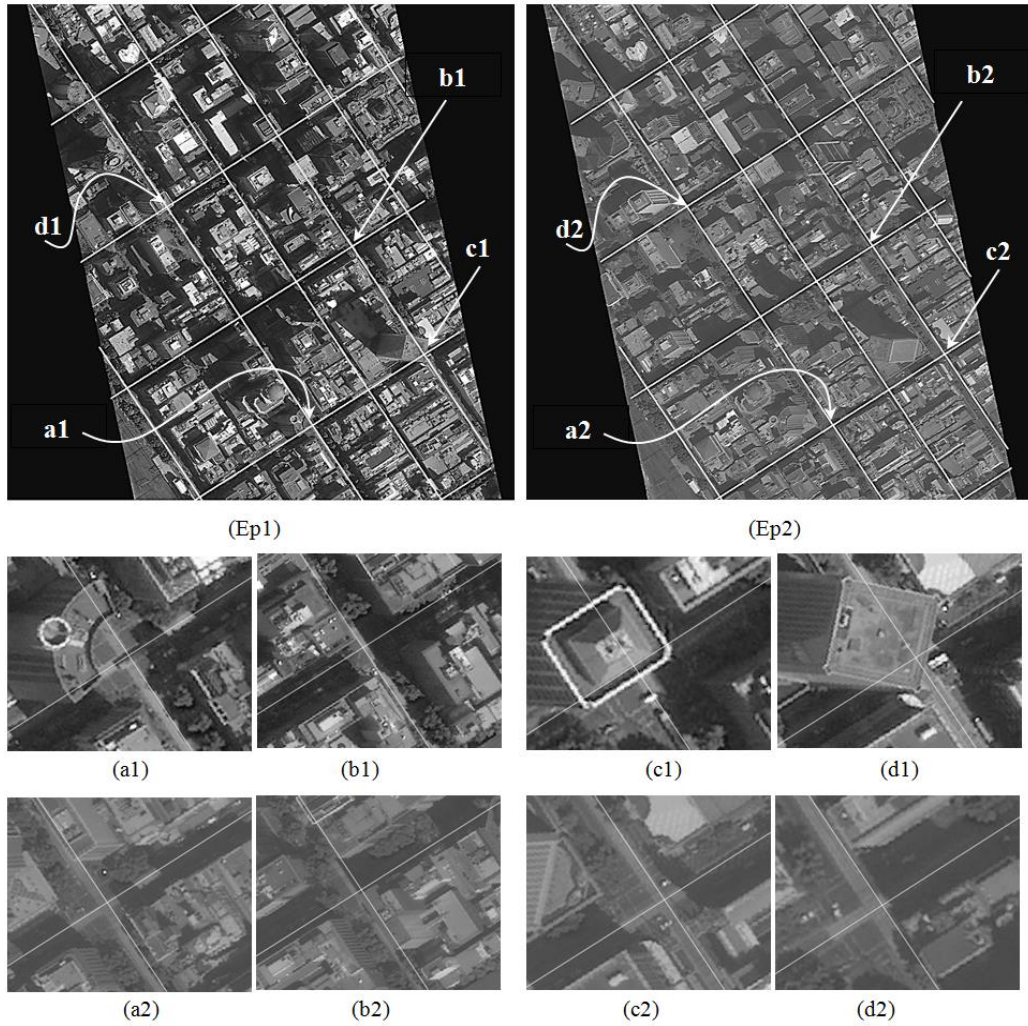
Figure 3.7 The result of co-registereing roads center lines (straigh white lines) in the right epipolar (Ep1) and the left epipolar (Ep2) images. The sub-images are a magnification of the four road intersections in the left (a1, b1, c1, d1) and in the right (a2, b2, c2, d2) images. VHR satellite images.

The final step in the developed RMAD technique was generating the disparity map (**Figure 3.8**) by performing the SGM algorithm. The disparity map was an exact fit to the reference epipolar image (i.e., the left image). It is noticeable that the straight-line roads have a disparity of almost zero value (black pixels between building edges in **Figure 3.8-(b)**). This is due to the success in the co-registration of the terrain roads in both stereo

81

epipolar images. The measured non-zero disparity values represent almost all the off-terrain objects including buildings and trees.

<div align="center">(a)   (b)</div>
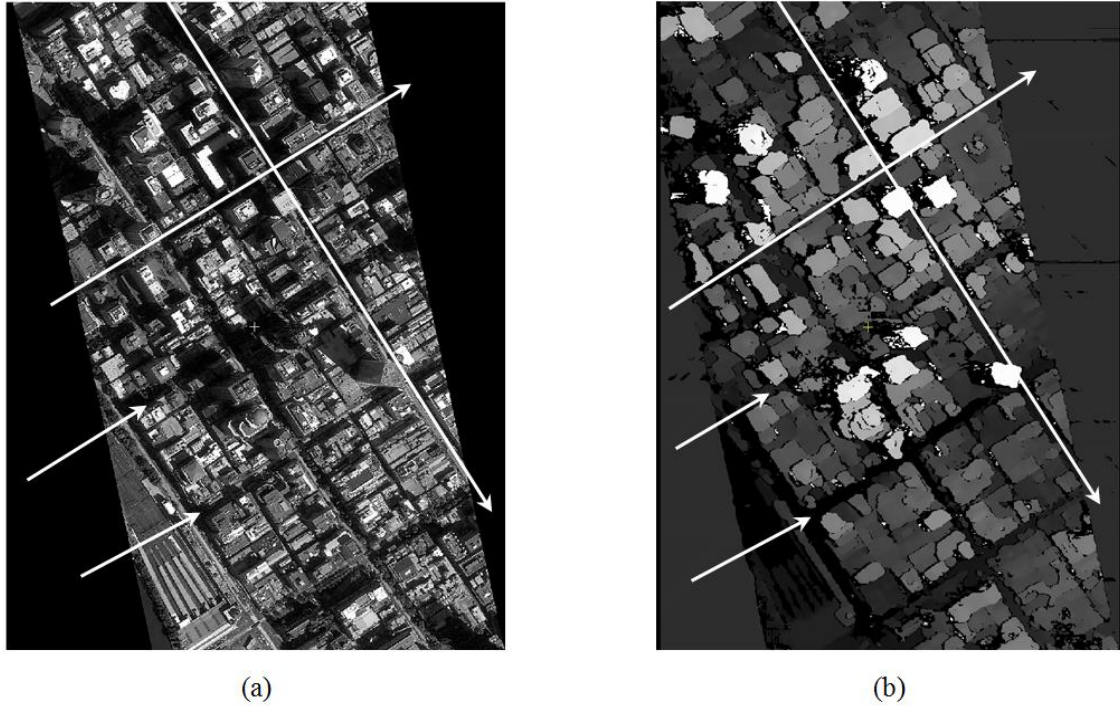
Figure 3.8 Aboveground disparity map. (a) The left VHR epipolar image, and (b) the resulting aboveground disparity map generated by RMAD technique using SGM algorithm.

3.4.2.2   Results of building detection based on RMAD technique

The building detection algorithm developed in **Section 3.2.2** was applied on the study area. The VHR epipolar imagery was over-segmented using a small scale (≈ 60). After calculating the vegetation index (NDVI), all trees were removed from the detection result using a threshold value close to 0.2, as recommended by many researchers (Suliman & Zhang, 2015). Then, all off-terrain building segments were detected based on the positive disparity values (i.e., > zero). Finally, building segments were merged

together, and the segments of small areas were filtered out based on empirically selected threshold value.

Figure 9 provides the disparity map, the detected building roofs, and the reference data for visual assessment of the quality of the proposed building detection framework. Comparing the detection result (**Figure 3.9-(b)**) and the reference data (**Figure 3.9-(c)**), it can be seen that almost all building roofs have been correctly detected. Furthermore, the shape and size of the most building roofs are almost the same as those in the reference data.

(a)   (b)   (c)
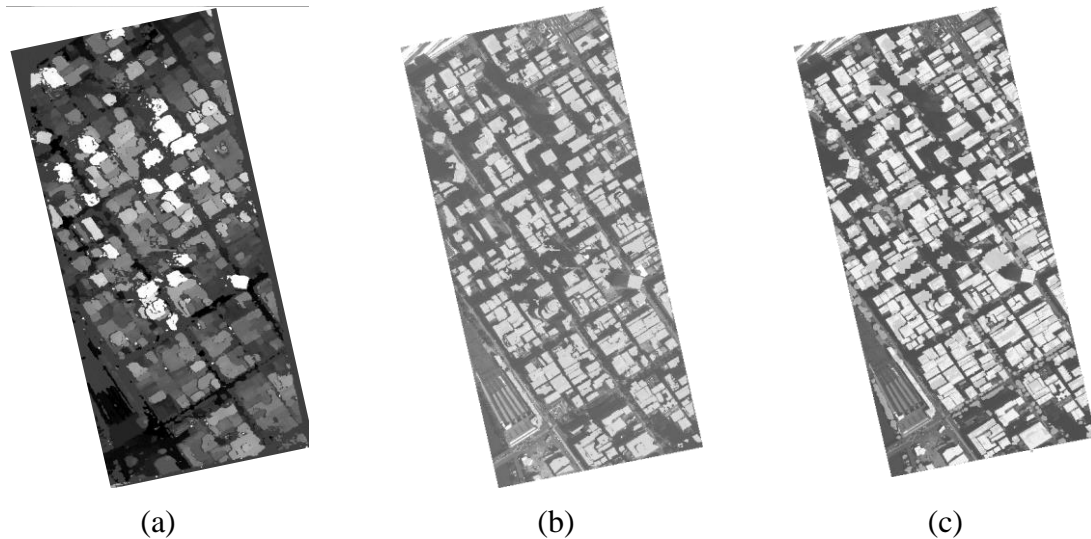
Figure 3.9 Comparison between detected building roofs and reference data building. (a) Aboveground disparity map generated based on RMAD technique. (b) Detected building roofs based on the aboveground disparity information. (c) The reference data for building roofs.

The three accuracy measures, as described in **Section 3.3.1**, are calculated for evaluating the detection performance of the proposed RMAD technique (**Table 3.1**). It can be seen from **Table 3.1** that three-fourth of the building roofs in the reference data are detected. This roof delineation result is correct by almost 100%. This reliable

detection is attributed to the precise disparity information used which led to overall detection quality of 75%. This quality is highly acceptable considering that the employed images suffer from a lot of occluded areas and that no other images or data were used.

Table 3.1 Accuracy measures for building detection using RMAD technique

Evaluation measures	Comp. (%)	Corr. (%)	OQ. (%)
Values	74	98	75

3.4.3 Results of Comparing Conventional Techniques

3.4.3.1 Results of disparity map generation and normalization using conventional techniques

The approach developed by Zhang et al. (2004) represents an alternative approach to the co-registration step of the RMAD technique. Their technique applies an interpolation step to fill the occluded areas and gaps in the surface representation, followed by extracting the underlying surface using a local minima algorithm. This extraction algorithm and the interpolation were applied to the generation of a disparity map using the same dataset. The critical input value required by this algorithm is the size of the moving window. A constant window size bigger than the largest building roof in the scene was selected based on the approach developers' recommendation and a few trial implementations of our test. The results of the individual steps of this technique are illustrated in **Figure 3.10**.

Figure 3.10 shows that the filled gaps by interpolating the surrounding data lead to misleading information that destroy the quality of the generated disparity maps. This

inaccurate information affects the normalization process negatively when the extracted terrain information is subtracted from the interpolated disparity map. This negative effect will reduce directly the accuracy of the building detection.

<div align="center">(a)   (b)   (c)   (d)</div>
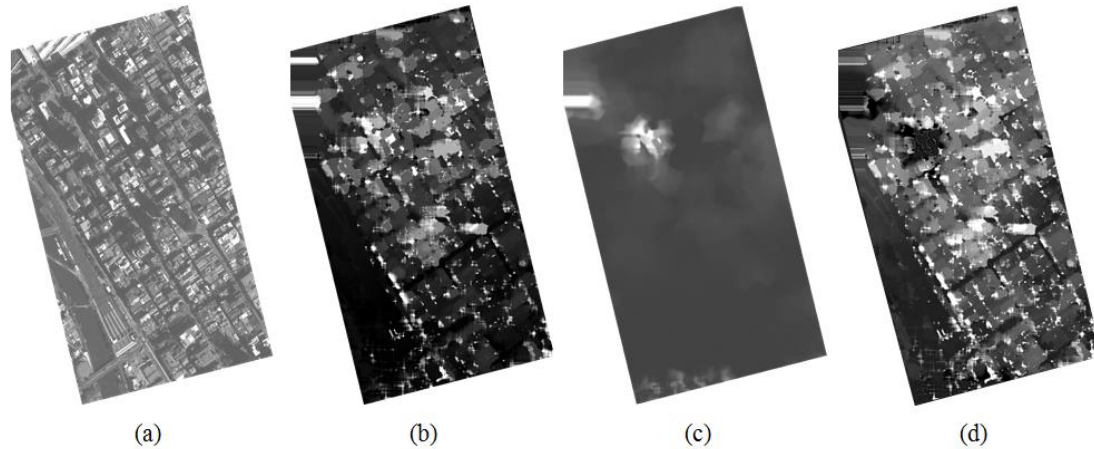
Figure 3.10 Disparity maps generated through conventional interpolation and normalization processes. (a) The left epipolar VHR image. (b) The disparity map after gap-filling interpolation. (c) Terrain disparity extracted by local minima technique. (d) Normalized disparity achieved by subtracting (c) from (b).

3.4.3.2 Results of building detection based on conventional techniques

The disparity-based building detection procedure developed by the authors and described in **Section 3.2.2** was implemented using the output from the previous steps. Specifically, the epipolar images were generated based on Zhao et al. (2008) and the normalized disparity map was generated following the approach of Zhang et al. (2004). The building rooftops detected based on our approach are shown in **Figure 3.11**, which indicates that the achieved result is of low quality. The false positive detection of building rooftops, shown in light gray in **Figure 3.11-(b)**, are mainly on the narrow roads and between buildings. The accuracy evaluation measures for this result are provided in **Table 3.2**.

Table 3.2 Accuracy measures for building detection based on published techniques for epipolar image generation and surface data normalization.

Evaluation measures	Comp. (%)	Corr. (%)	OQ. (%)
Values	82	48	43

☐ Dark gray: true positive (TP)
☐ Light gray: false positive (FP)

(a) (b) (c)
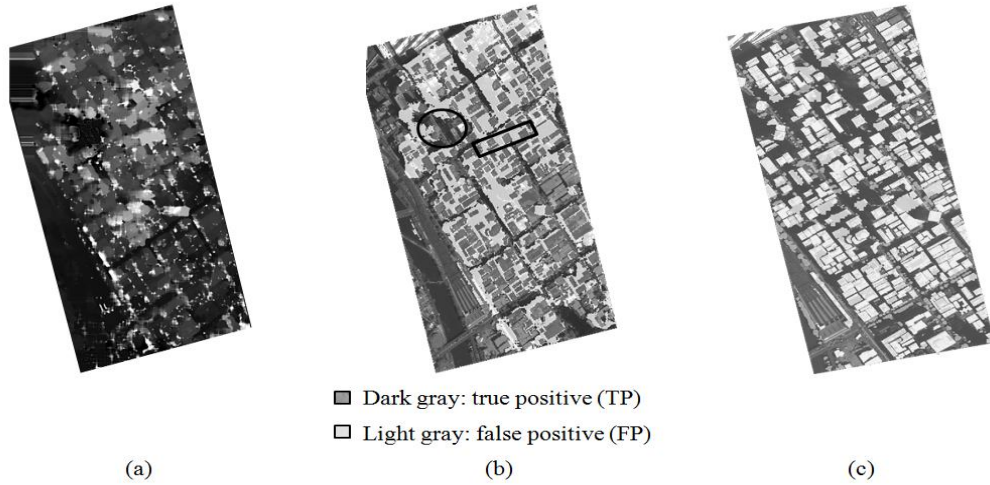
Figure 3.11 Results of building detection using conventional techniques. (a) Aboveground disparity map generated using RFM-based epipolarity model and normalized using local-minima-based technique. (b) Detected building roofs. (c) Reference dataset.

Figure 3.12 depicts two major negative outcomes identified in dense urban areas. In this respect, two specific areas marked in **Figure 3.11-(b)** are magnified in the optical, disparity, and mapping domains respectively in **Figure 3.12**. The upper row of **Figure 3.12** represents the effect of interpolating the resulting gaps in the disparity map generated in dense urban areas that led to false positive detection shown in light gray. The lower row represents the effect of normalizing the surface disparity resulting from subtracting the extracted terrain disparity model which led to missing building roofs.

Optical Domain	Disparity Domain	Mapping Domain
The optical VHR image	Interpolated disparity (misleading SDM)	Misdetection due to the misleading nSDM
The optical VHR image	Normalized disparity (misleading nSDM)	Missed roofs due to the misleading nSDM
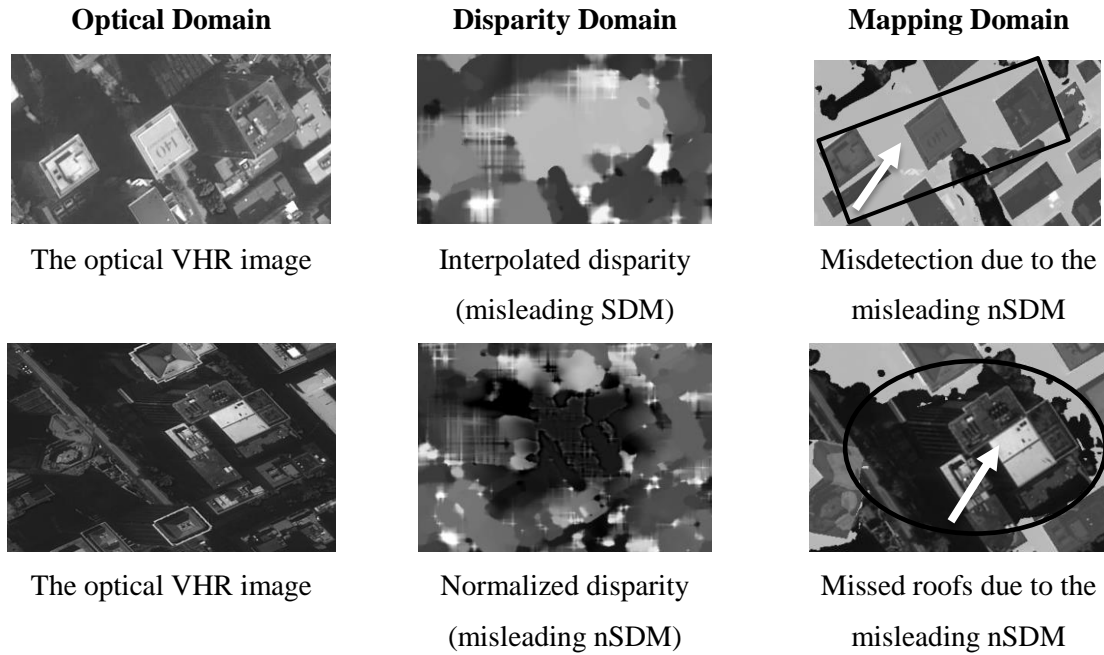
Figure 3.12 The two identified interpolation problems that create misleading information.

3.5 Results Comparison and Discussion

3.5.1 Results of the Achieved Aboveground Disparity Maps

A comparison between the aboveground disparity maps generated based on the developed RMAD technique versus the normalized map achieved using the conventional techniques is illustrated in **Figure 3.13**. The RMAD-based disparity map (**Figure 3.13-(a)**) shows that the disparity values of the narrow ground-level roads are almost eliminated and the adjacent building roofs are isolated even without trees suppression. Conversely, when the conventional normalization method (local-minima-based technique) is used, most of the adjacent buildings are merged and the narrow streets are

missed (**Figure 3.13-(b)**). This is attributed to the effects of gaps-filling by interpolation in crowded dense urban area.

<div align="center">(a)   (b)</div>
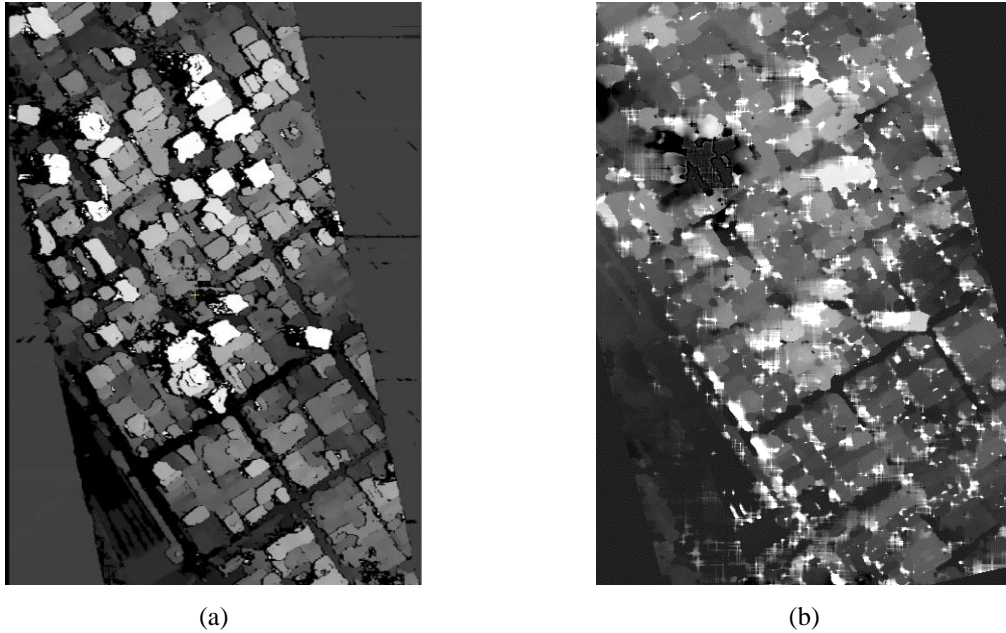
Figure 3.13 Comparison between the achieved aboveground disparity map generated by (a) RMAD, and (b) the surface disparity map (nSDM) normalized by local-minima-based technique.

The RMAD technique normalizes the disparity information through co-registering the ground-level features in both epipolar stereo images. The result, shown in **Figure 3.7**, is of acceptable quality based on the visual assessment of the straight-line streets. The main factor for this successful registration is the minimization of the reasonable terrain variation by the image rectification step using the average terrain variation. This rectification eliminated almost all the disparity of ground-level features as shown by the black pixels (zero-disparity) in the main and narrow streets in the study area (**Figure 3.8** and **Figure 3.13-(a)**). The disparity map generated in this figure represented almost all the aboveground objects.

On the other hand, the disparity map generated based on the RFM-based epipolarity technique needed to be normalized. The RFM-based technique does not rectify or minimize terrain relief distortion, the outcome of which caused terrain features to have disparity variation. Additionally, the need to fill the gaps by interpolation resulted in further deterioration of the generated surface disparity in dense urban areas (SDM in **Figure 3.10-(b)**). Continuing with the approach by Zhao et al. (2008), the TDM was extracted by the local-minima technique (**Figure 3.10-(c)**) which contributed to creating misleading information when subtracted from the interpolated SDM. The compound negative effects of the interpolation and local-minima filtering caused the omission of some buildings' disparity in the obtained nSDM (**Figure 3.10-(d)**).

3.5.2   Comparison of the Different Building Detection Results

For the ease of comparison, Figure 14 illustrates the results of the detected buildings using disparity information achieved based on both the developed RMAD technique and the conventional local-minima-based normalization technique. In the figure, the true positive and the false positive results are represented in dark and light gray respectively.

By visual inspection of the detection results in **Figure 3.14**, one can see the huge improvement in RMAD-based result relative to that achieved by the conventional technique. While the detected roofs in **Figure 3.14-(a)** are clearly separated from each other, these building roofs in **Figure 3.14-(b)** are merged together along with the narrow streets between. The quantitative analysis of these building detection results with the reference to the evaluation measures (Correctness, Completeness, and Overall Quality) are compiled from **Table 3.1** and **Table 3.2** and listed in **Table 3.3**.

89

(a)   (b)

Dark Gray: True Positive (TP)
Light Gray: False positive (FP)
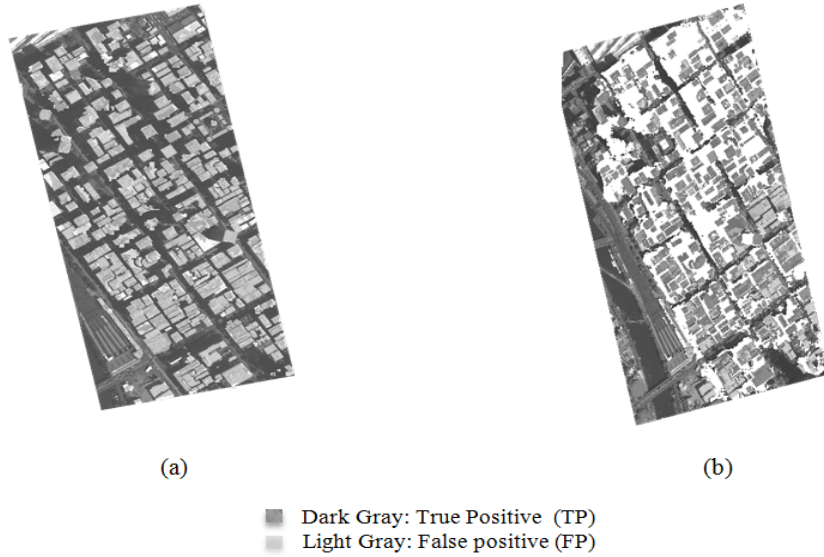
Figure 3.14 Comparison between the results of disparity-based building detection using (a) RMAD technique, and (b) using the normalized disparity map achieved by local-minima-based technique.

Table 3.3 Comparison of accuracy measures for building detection results achieved by RMAD technique and conventional normalization technique.

Building detection results based on	Corr. (%)	Comp. (%)	OQ. (%)
RMAD Technique	98	74	75
Conventional Technique	48	82	43

From the Correctness measures listed in **Table 3.3**, it is clear that RMAD-based procedure was able to provide correct detection of almost 100%. This is due to the reliable key feature used for the detection, which is the generated disparity information. It is worth mentioning that the detection was achieved by thresholding the generated disparity map by an almost zero-value. This further proves that the performed co-registration of the epipolar stereo images properly approximated the aboveground disparity map (i.e., nSDM). On the other hand, the conventional normalization technique produces a low Correctness value of 48% which indicates that the techniques for epipolar

90

generation and disparity map normalization are inefficient in detecting building's rooftops. The main reason for that is the misleading information resulting from the interpolation step to fill the gaps in dense urban areas. This incorrect information deteriorated the quality of the generated surface disparity map and, thereafter, the subsequent steps.

Regarding the Completeness measures provided in **Table 3.3**, about 25% of the building roofs in the scene were missed in the RMAD-based detection result. However, visual inspection of **Figure 3.9** indicates that the results were very close to the reference data. The main reason for misidentifying some rooftops is the lack of disparity information at these rooftops because they are occluded by higher elevated building roofs in one image of the stereo pair (the effect of this occlusion is matching failure and hence no disparity information will be available). Pixels in this category as well as misidentified pixels within and around a roof object were not included in the true positive count notwithstanding that the rooftop was correctly detected as a whole. Accordingly, the completeness measure decreased and quantitatively pointed out differently than the visual inspection.

On the other hand, almost one-fifth of the detection result was missed for the results generated based on the conventional interpolation and normalization techniques. Interpolating the gaps of the resulting disparity map in dense urban areas leads to smoothing the sharp building edges and missing the narrow streets between buildings. The consequence is that many terrain-level objects will be elevated falsely as buildings. Thus, many non-building objects were detected and classified as buildings as indicated by the high value of the Completeness measure in **Table 3.3**. The effect of the

interpolation on the detection result is illustrated in the upper row of **Figure 3.12**, where three high-rise and separated building roofs were mistakenly classified as one building object.

Data normalization by subtracting the extracted terrain disparity from the original surface disparity map amplified the effects of surface interpolation and filtration by local-minima technique. The result was an inaccurate TDM that led to missing actual building objects. The effect of normalizing the disparity map on the detection outcome is illustrated in the lower row of **Figure 3.12**. As shown in this figure, a few gaps on a street between two high-rise buildings were falsely filled and elevated by the interpolation process. Thus the local-minima technique extracted incorrect ground-level information and produced a highly elevated terrain area. This area was erroneously removed in the subtraction process and, consequently, no building roofs detected in this area of the nSDM.

The evaluation measure that combines both the Correctness and Completeness measures is the Overall Quality (OQ) of the detection. The RMAD-based detection overall quality, of 75% as in **Table 3.3**, is affected directly by the missed roofs from the detection. However, the achieved quality is still promising for building roof detection using only disparity information. Moreover, there are no building façades detected in the results due to having VHR stereo images of opposite backward and forward viewing angles as discussed in **Section 3.2.2**.

In contrast, the Overall Quality of the detection performance achieved based on the conventional interpolation and normalization techniques was only 43%. This low value was affected directly by the high false Completeness and low Correctness values. This

low value proves the severe negative effects of the interpolation process in dense urban areas. In contrast, the developed RMAD technique provided an efficient solution for such dense areas. The degree of the improvement in the tested area in this study was 32%. Obviously this value depends mainly on the nature of the urban area. However, it shows an attractive degree of improvement for the proposed disparity normalization technique based on co-registration.

3.6 Conclusions

This paper developed a registration-based technique for disparity mapping of aboveground objects (RMAD) based on image rectification, epipolar-pair generation, and co-registration methods. The aim is to mitigate the impact of occlusion on surface representation. The developed technique achieved this goal by eliminating the need for interpolating the gaps in surface elevations and producing a direct approximation for the off-terrain disparity information. The key novel concept in the developed technique is the co-registration of terrain-level objects in both stereo VHR images. Consequently, the disparity measured for ground objects in the epipolar direction becomes very close to zero value. This directly approximates the normalized disparity information which could be easily incorporated in disparity-based building detection applications.

The implementation and testing of the approach proved its effectiveness in dense urban areas characterized by reasonable terrain variation. The terrain relief distortions are dramatically reduced after rectifying the images using the area average elevation. This allows acceptable co-registration for the terrain-level objects (e.g., roads).

When the generated disparity map was tested in a disparity-based building detection application, the Overall Quality of the achieved result was 75%. The Correctness of the detection was almost 100%, which emphasizes the reliability of the disparity information in detecting buildings. The buildings were detected by applying a close-to-zero threshold. This value further indicates the success of the proposed co-registration method in approximating the normalized surface disparity. There was no building façades wrongly detected, taking advantage of the opposite off-nadir viewing angles of the VHR line sensors. Additionally, the RMAD-based detection of building rooftops requires only one stereo-pair of VHR satellite images. This is advantageous as it reduces the cost of having multiple overlapped images to a minimum.

The developed RMAD technique is evaluated against the disparity-based building detection result derived using published techniques for epipolar generation (RFM-epipolarity technique) and disparity map normalization (based on the local-minima technique). The achieved detection result by these techniques was substantially lower than that using the RMAD technique. The main reason for that is the severe negative effect of interpolating over the occluded areas in off-nadir stereo pairs. Such situation produces misleading information that negatively impacts the quality of the disparity maps and all subsequent processes including the building detection.

All these attractive features and improvements prove the usefulness of the developed technique for disparity-based building detection in dense urban areas of reasonable terrain level variations. Additionally, it increases the usability of the off-nadir stereo images in urban mapping and modeling applications.

The future work of this research will investigate overcoming probable limitations to the developed method. The main one is expected to appear when the imaged area is of a hilly terrain. Defining the projection reference plane by the average terrain elevation for the whole imaged area (full scenes) may not be enough. In this situation, and for scenes larger than our test image, one may adopt a tile-based approach for implementing RMAD method to co-register terrain level features. Another challenge to the RMAD method is its use of NDVI for removing trees and vegetation.

ACKNOWLEDGEMENTS

REFERENCES

AASHTO. (2001). *A policy on geometric design of highways and streets 2001*. Washington, D.C.: The Association.

Alobeid, A., Jacobsen, K., & Heipke, C. (2010). Comparison of matching algorithms for DSM generation in urban areas from Ikonos imagery. *Photogrammetric Engineering and Remote Sensing, 76*(9), 1041-1050.

Arefi, H., Engels, J., Hahn, M., & Mayer, H. (2007). Automatic DTM generation from laser-scanning data in residential hilly areas. *International Archives of Photogrammtry, Remote Sensing, and Spatial Information Sciences*, *36*(W45), Stuttgart, Germany.

Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. And Griesbner,G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*, (pp. 12-23). Wichmann Verlag, Karlsruhe.

Beumier, C. (2008). Building detection in ikonos images from disparity of edges. *Proceedings of the 3ʳᵈ International Conference on Computer Vision Theory and Applications*, Funchal, Madeira, Portugal.

Chen, D., Shang, S., & Wu, C. (2014). Shadow-based building detection and segmentation in high-resolution remote sensing image. *Journal of Multimedia, 9*(1), 181-188.

Dey, V. (2013). Image segmentation techniques for urban land cover segmentation of vhr imagery: recent developments and future prospects. *International Journal of Geoinformatics, 9*(4), 15-35.

Ghaffarian, S., & Ghaffarian, S. (2014). Automatic building detection based on supervised classification using high resolution google earth images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(3), 101.

Habib, A., Morgan, M., Jeong, S., & Kim, K. (2005). Analysis of epipolar geometry in linear array scanner scenes. *The Photogrammetric Record, 20*(109), 27-47. doi: 10.1111/j.1477-9730.2005.00303.x

Heinl, M., Walde, J., Tappeiner, G., & Tappeiner, U. (2009). Classifiers vs. input variables—the drivers in image classification for land cover mapping. *International Journal of Applied Earth Observation and Geoinformation, 11*(6), 423-430.

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(2), 328-341. doi: 10.1109/TPAMI.2007.1166

Khosravi, I., Momeni, M., & Rahnemoonfar, M. (2014). Performance evaluation of object-based and pixel-based building detection algorithms from very high spatial resolution imagery. *Photogrammetric Engineering and Remote Sensing, 80*(6), 519-528. doi: 10.14358/PERS.80.6.519

Krauß, T., d'Angelo, P., Kuschk, G., Tian, J., & Partovi, T. (2015). 3D-information fusion from very high resolution satellite sensors. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(7), 651

Krauß, T., & Reinartz, P. (2010). Urban object detection using a fusion approach of dense urban digital surface models and VHR optical satellite stereo data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(1/W17).

Krauß, T., Sirmacek, B., Arefi, H., & Reinartz, P. (2012). Fusing stereo and multispectral data from WorldView-2 for urban modeling. *The SPIE 8390, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery*, Baltimore, Maryland, USA.

Nielsen, M., & Ahlqvist, O. (2014). Classification of different urban categories corresponding to the strategic spatial level of urban planning and management

using a SPOT4 scene. *Journal of Spatial Science, 60*(1), 99-117. doi: 10.1080/14498596.2014.943309

Oriot, H. (2003). Statistical snakes for building extraction from stereoscopic aerial images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 34*(3/W8), 65-72.

Rottensteiner, F., Trinder, J., Clode, S., & Kubik, K. (2005). Using the dempster–shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion, 6*(4), 283-300. doi: 10.1016/j.inffus.2004.06.004

Rutzinger, M., Rottensteiner, F., & Pfeifer, N. (2009). A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2*(1), 11-20.

Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of Geoinformatics, 8*(2), 35-51.

Singh, S., Jain, K., & Mandla, V. (2014). Image based 3D city modeling: comparative study. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40*(5), 537-546

Story, M., & Congalton, R. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing, 52*(3), 397-399.

Sugg, Z., Finke, T., Goodrich, D., Susan, M., & Yool, S. (2014). Mapping impervious surfaces using object-oriented classification in a semiarid urban region. *Photogrammetric Engineering and Remote Sensing, 80*(4), 343-352. doi: 10.14358/PERS.80.4.343

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016). A novel technique for mapping the disparity of off-terrain objects. *Proceedings of IOP Conference Series: Earth and Environmental Science, 34*(1), 012035.

Suliman, A., & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 1913-1923.

Tao, C., & Hu, Y. (2001). A comprehensive study of the rational function model for photogrammetric processing. *Photogrammetric Engineering and Remote Sensing, 67*(12), 1347-1358.

Tian, J., Cui, S., & Reinartz, P. (2014). Building change detection based on satellite stereo imagery and digital surface models. *IEEE Transactions on Geoscience and Remote Sensing, 52*(1), 406-417. doi: 10.1109/TGRS.2013.2240692

Van de Voorde, T., De Genst, W., & Canters, F. (2007). Improving pixel-based VHR land-cover classifications of urban areas with post-classification techniques. *Photogrammetric Engineering and Remote Sensing, 73*(9), 1017.

Wang, M., Hu, F., & Li, J. (2011). Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 347-355. doi: 10.1016/j.isprsjprs.2011.01.002

Weidner, U., & Förstner, W. (1995). Towards automatic building extraction from high-resolution digital elevation models. *ISPRS journal of Photogrammetry and Remote Sensing, 50*(4), 38-49.

Xie, Y., Weng, A., & Weng, Q. (2015). Population estimation of urban residential communities using remotely sensed morphologic data. *IEEE Geoscience and Remote Sensing Letters, 12*(5), 1111-1115. doi: 10.1109/LGRS.2014.2385597

Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering and Remote Sensing, 70*(6), 657-661.

Zhang, Y., Tao, C., & Mercer, J. (2004). An initial study on automatic reconstruction of ground DEMs from airborne IfSAR DSMs. *Photogrammetric Engineering and Remote Sensing, 70*(4), 427-438.

Zhao, D., Yuan, X., & Liu, X. (2008). Epipolar line generation from IKONOS imagery based on rational function model. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*(B4), 1293-1297.

Zhu, H., Li, Y., Liu, Z., & Fu, B. (2015). Estimating the population distribution in a county area in China based on impervious surfaces. *Photogrammetric Engineering and Remote Sensing, 81*(2), 155-163. doi: 10.14358/PERS.81.2.155

Chapter 4: DOUBLE PROJECTION PLANES METHOD FOR GENERATING ENRICHED DISPARITY MAPS FROM MULTI-VIEW STEREO OFF-NADIR VHR SATELLITE IMAGES [1]

Abstract

The use of disparity information is computationally cost-effective for 3D-assisted mapping applications. However, off-nadir satellite images over dense urban areas suffer from severe occlusion. Therefore, large gaps will be created due to the unsuccessful matching in the occluded areas. Commonly, gap interpolation produces misleading information that destroys the quality of the subsequent information extraction application. Hence, the more reliable solution is to fill the occlusion gaps by generating supplementary data. However, disparity maps, as per definition, are the co-relation of one stereo pair. Hence, supplementary disparity maps cannot be directly generated and applied. Thus, this paper introduces the Double Projection Plane (DPP) method for constructing multi-view epipolar satellite images that maintain the proportionality among the corresponding disparities measured from different stereo pairs. Additionally, based on two projection planes, the method calculates the transferring scales required for producing supplementary disparity data. Accordingly, based on multi-view stereo satellite images, this method is promising to provide gap-and-outlier-free disparity maps that have the potential to replace the need for elevation models to some extent.

4.1   Introduction

The currently available very high resolution (VHR) satellite sensors are able to provide along-track (inline) multi-view stereo (MVS) images of sub-metre ground resolution, at low cost, and with broad coverage. Digital elevation models (DEMs), orthophotos, and 3D rendered photo models are well-known products that can be reconstructed from stereo VHR satellite images based on extracting the third dimension. Such products are essential for many real-world applications including planning, designing, and management of urban environments. Consequently, stereo image processing of satellite images has become an active research area within the remote sensing community.

The extraction of the third dimension from stereo images is a well-established procedure in the field of Digital Photogrammetry. This procedure utilizes matching techniques to provide stereo-based image measurements known as disparities before they are used to derive object-space elevations. Hence, the third dimension may be represented by either the disparity or the elevation information. Since extracting the elevations photogrammetrically requires several calculation steps after measuring the disparity data from stereo images, the use of the disparity data instead is computationally cost-effective not only in the runtime but also in the number of computational steps. Furthermore, disparity data are adequate for most 3D-assisted information extraction applications.

Occlusion areas remain a standing problem for all stereo matching techniques since these areas are usually visible in only one image of the stereo pair (i.e., half-occlusion). Consequently, the generated disparity maps from stereo pairs usually suffer from two major challenges: failed and false matches. The failed matches are the unmatched pixels

that create no-data regions (gaps). On the other hand, the false matches are the incorrectly matched points that create outliers. These outliers are typically detected and removed from the disparity map. Therefore, both of these unsuccessful matches (failed and false matches) produce gaps that need to be filled.

In the case of matching stereo off-nadir satellite images acquired over dense urban areas, the two challenges and the resulting gaps are further exacerbated due to the excessive building lean. This lean usually hides large areas in one image of the stereo pair leading to unsuccessful matching. Moreover, different building façades may have a similar appearance that confuses matching algorithms and hence creates additional false matches. Thus, disparity maps generated from stereo off-nadir images over dense urban areas usually suffer from large occlusion gaps. **Figure 4.1** shows a case of stereo images that contain an occluded area and a confusing similarity that are usually encountered in VHR stereo images when acquired off-nadir over urban areas. This pair will produce a large number of failed and false matches.
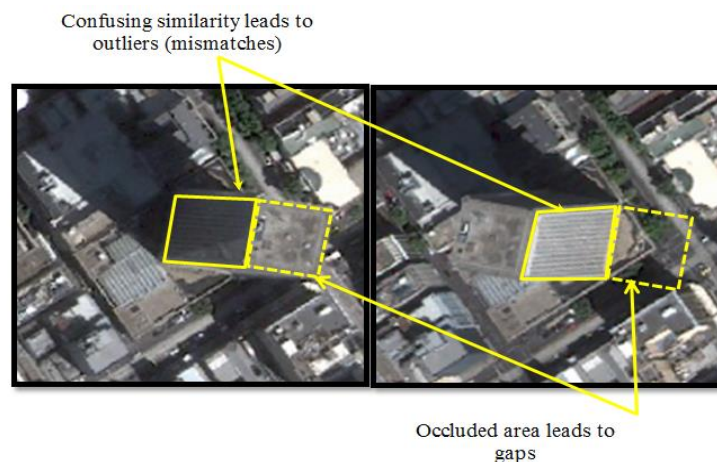
Figure 4.1 Stereo off-nadir VHR satellite images over an urban area where building lean occludes some areas and introduces confusing façade similarity.

The typical solution to fill the resulting disparity gaps is through applying interpolation techniques. However, executing these techniques in dense urban areas creates misleading information (such as missing narrow streets between buildings) and hence destroys the quality of the measured data. In contrast, filling the disparity gaps by integrating additional disparity data generated from other stereo pairs (referred to as supplementary disparity data in this paper) is a more reliable solution. Unfortunately, supplementary disparity data are usually not inter-proportional (scale-related) among different stereo pairs. This follows from the fact that a disparity value is the shift (in pixels) between two stereo epipolar-rectified images (i.e., image pairs with eliminated y-parallax). This means that the measured disparity maps represent a relative relation between only two stereo epipolar images along the epipolar direction. As a result, different and inconsistent disparity values will be measured for the same object-space point at different stereo pairs. This inconsistency poses a challenge for integrating disparity data from different stereo pairs.

The aim of this research is to develop a method to generate an enriched disparity map (i.e., significantly free of gaps and outliers) from multi-view inline stereo VHR satellite images. Crucial to this goal is achieving scale-related disparities, and deriving their transferring scales among different stereo pairs, in the object-space from multiple epipolar images that allow filling the occlusion gaps. Early work on developing the basis for generating scale-related disparities was briefly presented in Suliman, Zhang, and Al-Tahir (2016a). Extending that work, this study develops the method of double projection planes (DPP) to calculate the required disparity transferring scales.

The remainder of the paper is organized as follows: **Section 4.2** reviews the related work. The concept and the steps of the DPP method are presented in **Section 4.3**. The implementation and validation procedures are described in **Section 4.4**. The data, results, and the discussion are provided in **Section 4.5**. Finally, the conclusions are drawn in **Section 4.6**.

4.2   Literature Review

Unsuccessful matches reduce the quality of the generated disparity maps. These unsuccessful matches can be detected by visibility checks (e.g., the Left-Right check where the first image is matched with the second image and vice versa) that identify unstable matches (failed/false matches) which cause gaps and outliers. A comprehensive review of the most common techniques of detecting the visible areas in one image of the stereo pair is provided in Egnal and Wildes (2002). Commonly, once the occlusion has been detected, a gap filling process is then implemented. A few techniques to fill the gaps in disparity maps may be found in both the computer vision and remote sensing research communities. These techniques are based on either gap interpolation or supplementary data fusion.

4.2.1   Interpolation-based Techniques

In the field of computer vision, many interpolation-based algorithms have been proposed for occlusion filling in disparity and depth maps. Oh and Kuo (2010) interpolated the occlusion pixels from the non-occluded ones that have similar color intensity. Min and Sohn (2008) utilized the energies obtained from the matching process to fill the occlusion gaps by diffusing the energies of neighboring non-occluded points.

Hosni, Bleyer, Gelautz, and Rhemann (2009) filled the gaps of an occluded point from the minimum–to reflect the occluded background–among the disparities of its horizontally closest left and right non-occluded points. All of these algorithms apply restricted estimation and interpolation techniques to fill the disparity occlusions based on cues from the color intensity images. However, as reviewed by Huq, Koschan, and Abidi (2013), occlusion filling in these ways often introduces some streaks in the disparity maps. To minimize such streaks, Cho, Sun, Ha, and Jeong (2013) introduced a more conservative gap filling algorithm. They classified the disparity gaps into four cases that need to be manipulated differently. Nonetheless, in their general case, the occlusion gaps are filled by the minimum disparity of the left and right sides since only two stereo images are available.

On the other hand, remote sensing techniques tend to use additional knowledge from the optical imagery for gap-filling and enhancing the stereo-based height models (disparity maps and DEMs). The extracted image information is utilized in restricting the interpolation process for filling the data gaps. For example, Krauß and Reinartz (2010a) applied radiometric-based segmentation on a pan-sharpened VHR image. The generated image segments are used to restrict the gap interpolation within the boundary of each segment. Krauß and Reinartz (2010b) reduced the urban environment complexity by applying spectral classification to optical VHR imagery and dividing it into four main classes. These classes are manipulated differently based on some knowledge about each class to interpolate the data gaps. If the optical data were of high classification power (e.g., eight-band WorldView-2 optical stereo data), the same classification-based

refinement approach can be extended as implemented by Krauß, Sirmacek, Arefi, and Reinartz (2012).

Based on the conducted review, the alternatives for filling the gaps in the disparity maps are usually computationally expensive and exploit only one pair of stereo images. As a result, the disparity estimation and interpolation process are not independent from the generated disparity map. In the case of a pair of off-nadir stereo VHR images acquired over a complex urban environment, the performance of these filling techniques is further degraded by the presence of large occlusion gaps. Alternatively, our study proposes the generation of supplementary disparity data from multi-view stereo images that would be fused together for filling the gaps.

4.2.2   Fusion-based Techniques

In the computer vision field, generating supplementary data from MVS images has been an active research area. Many techniques have been introduced for both static and real-time applications, for example, Zhang, Cao, Zheng, Chen, and Wang (2014); Unger, Wahl, Sturm, and Ilic (2012); Merrell et al. (2007); Goesele, Curless, and Seitz (2006); Strecha, Fransens, and Van Gool (2006); and Szeliski (1999). In such studies, three basic steps are employed: (1) generating additional disparity maps (or depth maps) from the MVS images, (2) projecting/transferring the additional maps generated from different stereo pairs to the reference image of the reference stereo pair, and (3) fusing the projected maps to achieve accurate and dense representation. For transferring the supplementary disparity maps to the reference image of the reference stereo pair, a triangulation process is executed to calculate the 3D object-space location and then

followed by back projection of that location to the reference image for computing the corresponding disparity value in the image space. This step involves additional computations and adjustments.

In almost all of the reviewed computer vision literature, the MVS images used to generate supplementary data are close-range images acquired by frame sensors where each image has one perspective center. This sensor type has a relatively simple geometry that provides several advantages including (1) fixed, and usually known, baseline values for each stereo pair; (2) straight epipolar lines; and (3) direct scale relationship between the disparity and depth values. In contrast, remote sensing satellite images, as they are the focus in our research, have a more complex geometry. These images are acquired by line (push-broom) sensors and hence most of the advantages associated with frame sensors are lost. For instance, each scan line of the images acquired by line sensors has a different perspective center. This acquisition geometry yields non-straight epipolar lines when the images are rigorously epipolar rectified (Habib, Morgan, Jeong, & Kim, 2005).

In contrast to the above mentioned works, a successful technique of scaling and fusing disparity data is implemented by Hirschmüller (2008) for frame and push-broom sensors of non-space-borne images with low altitudes. The fusion process is based on transferring supplementary disparity values generated from different stereo images to the reference disparity map by scale values related to the lengths of the baselines between the matched stereo pairs. Hence, the technique requires the baseline length information to be available in advance. Additionally, this technique assumes that all the employed images are projected onto a common plane that has the same distance to all optical centres. This condition is challenging to meet for all image pixels when satellite images are used due to

the usual shifts and drifts in the satellite trajectory, and the perspective viewing of the satellite sensors in the direction perpendicular to the flight/scanning direction. Hence, pixel-based calculation for the transferring disparity scales in the common object-space projection plane should be considered when dealing with VHR remote sensing satellite images. This is to account for the factors related to the satellite trajectory instability, partial perspective viewing acquisition, and the baseline complexity (i.e., polynomial function instead of constant value).

In the remote sensing field, the vast majority of the studies are conducted based on elevation data. MVS images are used to generate supplementary elevation data that can be applied directly in a gap-filling process. This is because all of the stereo images are absolutely oriented and have the same object space. Theoretically, this leads to calculating the same elevation value for the same planimetric location regardless of which stereo pair is used to generate that elevation. However, this is not directly applicable for the disparity maps since they represent the relative relationship between only two images. This means that the disparity value measured for an image pixel with respect to one image will not be equal to or consistent with the same pixel disparity measured from any other stereo mate due to the change in the relative orientation. The disparity value, as a result, will not be directly transferable from one stereo domain to another.

Hence, studies that involve fusing supplementary disparity data for information extraction in the remote sensing community are extremely limited. In fact, no studies that exploit the merging of supplementary disparity maps could be found for the case of space-borne MVS images. In contrast, one successful study based on airborne MVS

images acquired by ADS line-scanner was conducted by Gehrke, Downey, Uebbing, Welter, and LaRocque (2012). They generated an information cloud by projecting both the measured disparity maps and the multispectral information into the object space. The supplementary disparity data are scaled and merged to enhance the reference disparity map. Unfortunately, the details of the scaling and merging processes were not provided.

Thus, the focus of our research is on developing scale-related supplementary disparity information from multiple inline stereo satellite images to be fused for filling the occlusion gaps in the reference disparity map. To solve the posed problem, the consistency property must exist in the epipolar rectified satellite stereo images. This property should be achievable if the satellite images are processed in the object space. However, because satellite sensors create curved epipolar lines, the relevant disparities from different stereo pairs will be inconsistent and hence not scale-related among different stereo domains. Therefore, for generating supplementary disparity data efficiently, we propose that the MVS-VHR satellite images should be epipolar rectified in a common object-space projection plane. A horizontal epipolar projection plane at the average terrain level of the imaged area allows both straightening the epipolar lines and equalizing the corresponding disparities of the same-elevation points (Morgan, Kim, Jeong, & Habib, 2006; Wang, Hu, & Li, 2011). Then, a second projection plane parallel to the epipolar rectification plane should be exploited to calculate the disparity transferring scales.

4.3 Double Projection Planes (DPP) Method

4.3.1 DPP Concept

The aim of the DPP method is to build inter-proportionality (scale-relation) among the disparities of different disparity maps and to calculate the transferring scales. The absolute orientation of the MVS-VHR satellite images in the relevant object space should create a direct relationship to achieve that aim. Hence, the MVS-VHR satellite images have to be epipolar-rectified and absolutely oriented in the object space. To analytically determine the scale values, a second projection onto the object space is required to calculate offset information that allows determining the disparity scales. Therefore, disparity values can be transferred among different stereo domains. This allows supplementary disparity data to be generated.

Disparity maps represent the relation of one epipolar image (reference image) with respect to another one (supplementary image) that both have a common overlap. Hence, the disparity value (D) is simply the X' coordinate value of a pixel in the domain image (usually the left stereo image, L) minus the X' coordinate value of the same pixel in the stereo mate (usually the right stereo image, R) as expressed in **Equation (4.1)**. The resulting disparity map fits exactly (i.e., shares the same reference frame) the reference epipolar image with pixel-level registration accuracy. However, this map usually has gaps that can be filled by transferring the corresponding disparities from other stereo pairs.

$$D_{LR} = X'_L - X'_R \qquad\qquad (4.1)$$

The concept of building and calculating the disparity transferring scales is illustrated in **Figure 4.2**. This figure, which is in the epipolar direction, shows three MVS-VHR satellite images that are projected on two different horizontal planes in the object space. While the first projection plane is at the elevation (Z_1), the second projection plane is at the elevation (Z_2). The light rays projected onto the two horizontal object-space planes are all represented in the epipolar plane, X' Z, along the same epipolar line. The intersection point of these light rays is the actual point imaged in the object space.

The intersections of the rays with the first horizontal plane are the object-space locations of the projected image pixels. Hence, the distance between these projected pixels in the epipolar direction are the object-space disparities (D). The moving distances ($\Delta X'$) of the projected image pixels in the second projection plane (Z_2) relative to the first one (Z_1) are the offset information required to derive the disparity transferring formula.

To derive the scales for transferring the disparities, basic triangle similarity rules can be applied. It can be found that two object-space pixels are enough to predict the location of the third one. This is achievable by having the offsets in the epipolar direction ($\Delta X'$) calculated for all pixels of the projected images. To account for the factors that may affect the accuracy of the calculated disparity transferring scales, the calculation is developed for each pixel of the projected images. These factors include the usual drifts in the satellite trajectory, the field of view of satellite sensors and the perspective viewing of the push-broom sensors in the direction perpendicular to the flight/scanning direction. **Equation (4.2)** describes the relationship derived for converting a disparity value between two domains (e.g., D_{nj} to D_{mn}) for each projected pixel (pixel-based calculation)

$$D_{m.n} = \left(\frac{\Delta X'_m - \Delta X'_n}{\Delta X'_n - \Delta X'_j} \right) . D_{nj} \qquad\qquad (4.2)$$
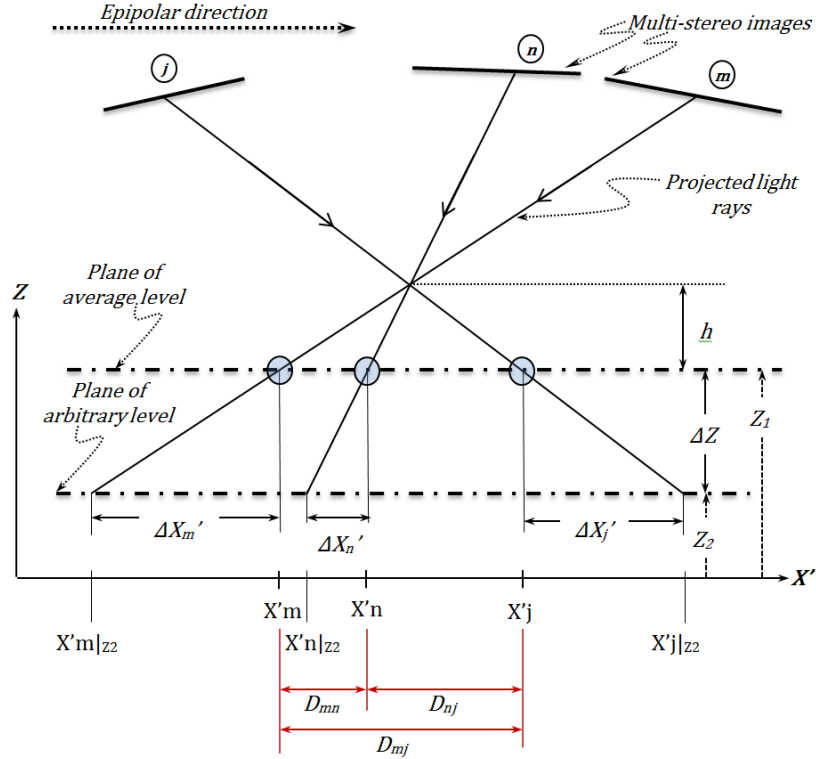
Figure 4.2 The concept of the proposed double projection planes (DPP) method

4.3.2 DPP Method

To implement the described concept, three steps as flowcharted in **Figure 4.3** are proposed for executing the DPP method. The epipolarity model introduced by Wang et al. (2011) is selected for developing the DPP method because it processes the satellite images in the object space. Also, this model has been proven to work accurately with full and long scenes without being affected by the relief displacement of the area even with the presence of mountains. In this epipolarity model, the sensor model information for all MVS-VHR satellite images must be available in advance.

111
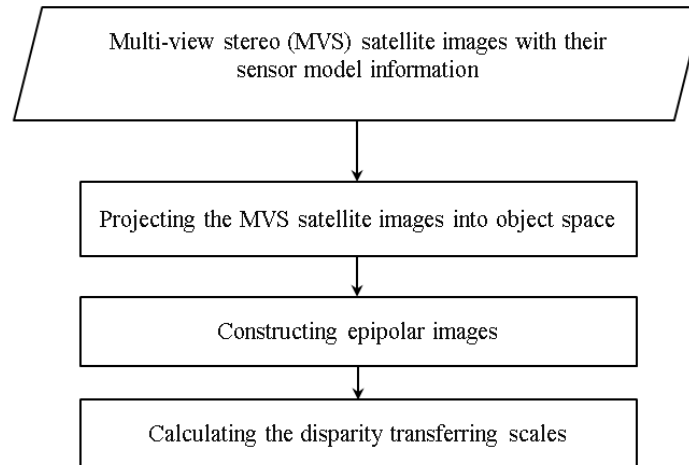
Figure 4.3 The proposed steps of double projection planes (DPP) method

For VHR satellite images, the rational function model (RFM) is the typical sensor model. It has two different sets of rational polynomial coefficients (RPCs) for two opposite transformation forms that establish the relationship between the object and image spaces. They are forward (ground-to-image) and backward/inverse (image-to-ground) transformation forms (Tao & Hu, 2002). Although the two sets of the RPCs are different and not correlated, most of the satellite image vendors provide the end users with one set of RPCs, namely the forward RFM model. If the RPCs of the inverse model are not available, they should be computed first. Readers who need to compute the backward RPCs are encouraged to consult the detailed explanations and calculations documented in Grodecki, Dial, and Lutes (2004) and Tao and Hu (2001).

4.3.2.1 Double projecting the satellite images into object space

The double projection process is based on two horizontal projection planes in the object space. It comprises two transformations based on the sensor model information of each MVS-VHR satellite image. The two planes and the two transformations are

112

illustrated in **Figure 4.4**.   The first transformation is to find the image location

corresponding to its object-space location at elevation Z_1. This transformation is based on

the forward transformation form. The second transformation is from the computed image

location back to the object-space location at elevation Z_2 to calculate the offset

information.
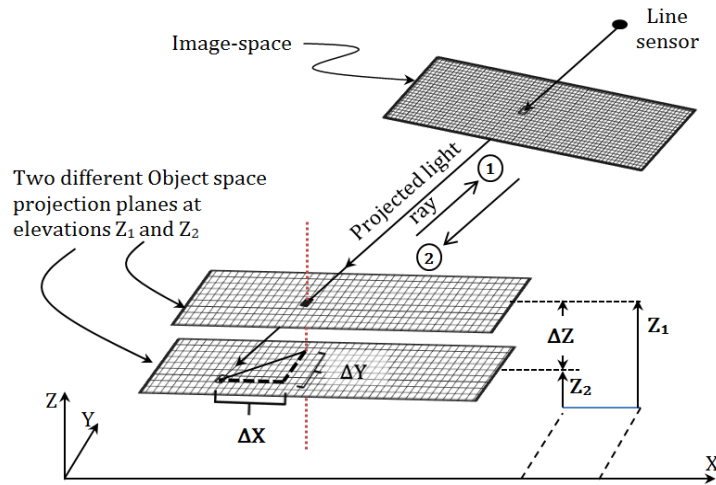
Figure 4.4 The proposed projection process at two object-space planes of different
elevations. An image pixel is projected at two different elevations (Z_1, and Z_2) to generate
two collinear 3D points for calculating its objects-space offsets. The offsets are used to
calculate the disparity transferring scales.

Regarding the first transformation, it is executed by (1) finding the approximate

ground coverage of the MVS images from the outer boundaries in their meta-data, (2)

constructing a 2D object-space grid with a spacing distance usually similar to the ground

resolution of the images, (3) specifying a constant elevation (Z_1) to represent a projection

plane parallel to the datum, and (4) projecting and resampling the MVS images to the

constructed object-space grid. The RFM forward transformation form, as described in

Tao and Hu (2002), should be used to find the corresponding image location.

The specific elevation value for the projection plane (Z_1) is the average terrain elevation of the imaged area in order to minimize the terrain relief distortions and to straighten the epipolar lines of the projected satellite images (Wang et al., 2011; Suliman, Zhang, & Al-Tahir, 2016b). The resampling process should be applied for all bands in the employed VHR images to allow subsequent phases of image analysis and information extraction.

The second transformation is executed using the RFM inverse transformation form based on the image location calculated in the first transformation and a different constant elevation (Z_2). This transformation calculates a 3D object-space point, at the elevation Z_2, which is collinear with its corresponding one at the elevation Z_1. This point location is required to determine the offsets of the object-space pixel location $(\Delta X, \Delta Y)$ in both X and Y directions as illustrated in **Figure 4.4**. The figure shows that the first transformation is from the constructed 2D object-space grid to the image space in order to resample the image brightness, and the second transformation is from the same image location to a collinear object-space point at different elevation (Z_2). Then, the offsets can be easily calculated as in **Equation (4.3)** for each image pixel (P_i)**.**

$$\Delta X|_{P_i} = X_2|_{P_i} - X_1|_{P_i} \quad , \quad \Delta Y|_{P_i} = Y_2|_{P_i} - Y_1|_{P_i} \tag{4.3}$$

where $\Delta X|_{P_i}$ and $\Delta Y|_{P_i}$ are the calculated object-space offsets for each pixel (P_i) in both X and Y directions respectively. $(X_1|_{P_i}, Y_1|_{P_i})$ and $(X_2|_{P_i}, Y_2|_{P_i})$ are the calculated object-space locations of the pixel (P_i) at the elevations Z_1, and Z_2 respectively. The two 3D points are of different levels and collinear.

4.3.2.2 Constructing the epipolar images

After projecting the MVS-VHR satellite images onto the projection plane as described in the previous step, the curved epipolar lines of linear sensor images are approximated to straight lines (Wang et al., 2011). Furthermore, because this plane is in the object space and parallel to the datum, the consistency property of the object-space disparities is preserved. This means that the disparity values of the same two points in different domains will be scale-related to each other all over the map as depicted in **Figure 4.5**. The figure shows two points of the same elevation imaged by MVS satellite scenes. The image-space distances of these two points are different for each space due to the relief displacement. After the described projection process, these distances are unified (i.e., become consistent) in the projected images. Hence, the object-space disparities of these images became equal for all points having the same ground elevation.

To perform the epipolar rectification of these MVS images, the projected images need to be relatively oriented along the epipolar direction to make the corresponding points lie on the same row in all of the projected stereo images. However, the relationship achieved after projecting the stereo images is required to be preserved. Thus, the plane coordinate axes of the object space (i.e., X, and Y axes) should be rotated by the angle of the epipolar direction in order to be parallel with the rows' direction.
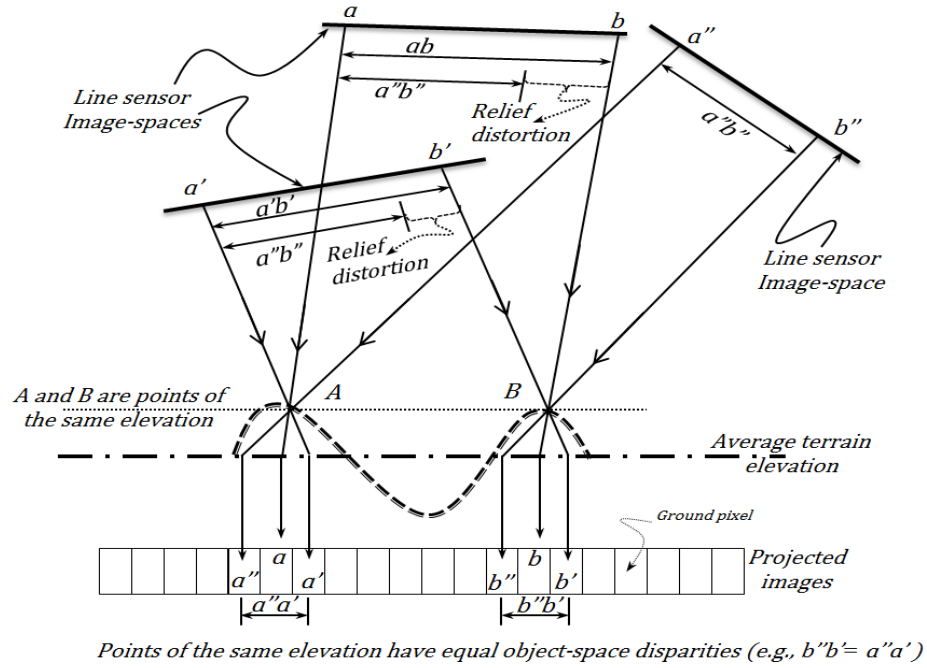
Figure 4.5 Object-space re-projection of the inline multi-view stereo (MVS) images.

The angle of the epipolar direction should be identified before applying the relative epipolar rectification. We introduce a direct equation to determine the epipolar direction, described in **Equation (4.4)**, based on the offset information calculated in the object-space (using **Equation (4.3)**) as illustrated in **Figure 4.6**. In this equation, any two of the projected MVS images can be selected to find two 2D offsets. All of these four offsets are for the same planimetric object-space location on the first projection plane.

$$\theta = \tan^{-1}\left(\frac{\Delta Y_a^n - \Delta Y_a^m}{\Delta X_a^n - \Delta X_a^m}\right) \tag{4.4}$$

where $(\Delta X_a^n, \Delta Y_a^n)$ and $(\Delta X_a^m, \Delta Y_a^m)$ are the two calculated 2D object-space offsets of a point (a) in the X and Y directions for two projected MVS images (n) and (m) respectively.
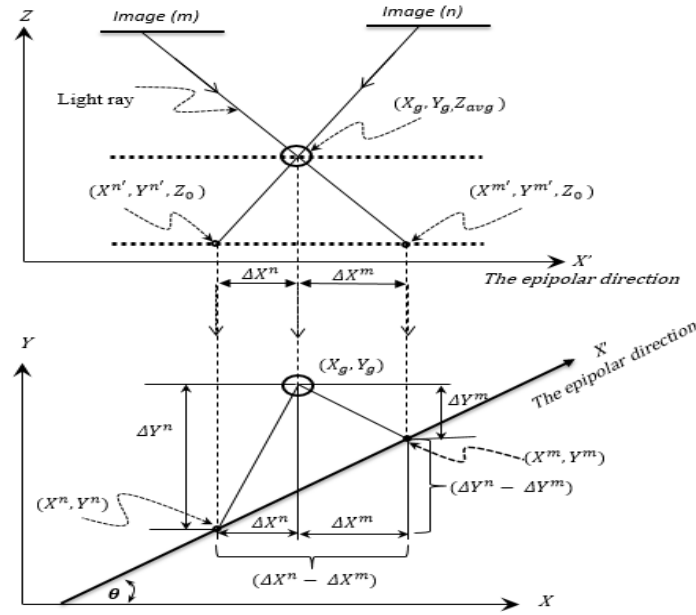
Figure 4.6 Deriving the epipolar direction from the calculated object-space offsets.

Figure 4.6 illustrates the derivation of the epipolar angle expressed in **Equation (4.4)**. It shows an object-space point, of the coordinates (X_g, Y_g, Z_{avg}), located within the overlapping area of all of the MVS images. This object-space point is shown in the figure in two views. While the upper part shows a plane along the epipolar direction ($X'Z$ plane), the lower part shows the top view of the object space ($X'Y'$ plane). The computed planimetric offsets of this object-space point for any two stereo images, say image (m) and (n), are $(\Delta X^m, \Delta Y^m)$ and $(\Delta X^n, \Delta Y^n)$ respectively. These planimetric offsets can be directly used to calculate analytically the angle (θ) as in **Equation (4.4)** which represents the epipolar direction in the object space.

A rotation by the amount of angle θ is then applied as described in **Figure 4.7** followed by resampling the projected images to make their rows parallel to construct the epipolar images and keep the relative orientation. The product of this step is a set of

117

epipolar images of the same size. This size is the minimum bounding box (*MBB*) that encloses all of the projected images.
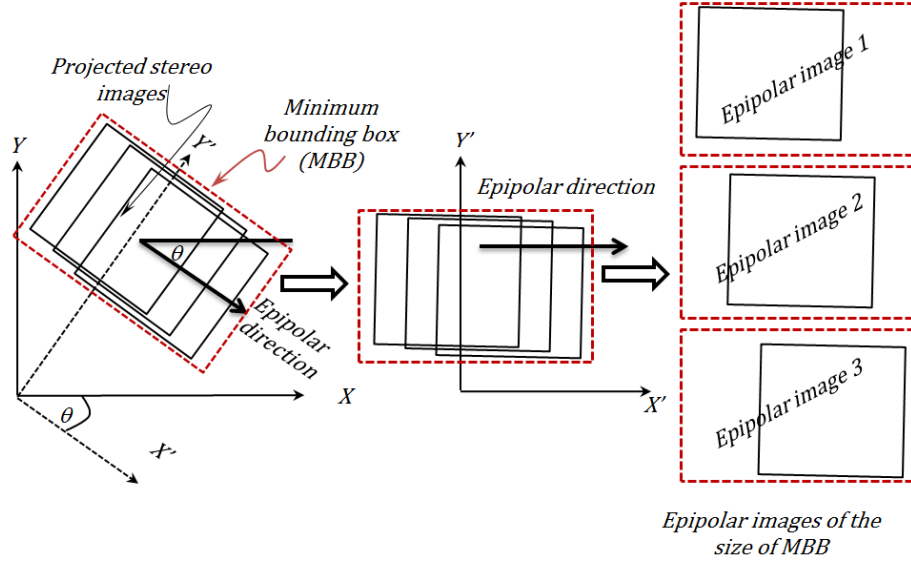
Figure 4.7 Epipolar image reconstruction.

4.3.2.3 Calculating the disparity transferring scales

Before calculating the disparity transferring scales, the offset information in the epipolar direction must be available. Hence, the computed offsets in the object-space coordinate system (i.e., ΔX, and ΔY) are needed to be rotated to calculate their corresponding offsets with respect to the epipolar direction (i.e., $\Delta X'$, and $\Delta Y'$).

To calculate the offset information in the epipolar direction, the rotation applied to the projected images, using the angle calculated in **Equation (4.4)**, should be applied again to the offsets (ΔX, and ΔY) calculated in the first step, using **Equation (4.3)**, for all object-space pixels of each image. This offset rotation can be achieved directly using **Equation (4.5)** which provides the offset information in two perpendicular directions. While the resulting offsets in the epipolar direction ($\Delta X'$) allow calculating the disparity

transferring scales, the perpendicular offsets ($\Delta Y'$) allow confirming the validity of the calculated offsets ($\Delta X'$) and the epipolar direction (θ) as well.

$$\begin{pmatrix} \Delta X' \\ \Delta Y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} \tag{4.5}$$

Once the offsets in the epipolar direction ($\Delta X'$) are determined for all projected images, the disparity transfer formula derived and expressed in **Equation (4.2)** can be applied. This formula allows transferring the disparity data among different stereo domains based on scale relationship. Hence, additional disparity data can be generated.

4.4   Methodology

The methodology for this study comprises both implementation and validation procedures. While the implementation procedure employs the DPP method to generate an enriched disparity map, the validation procedure assesses the quality of the achieved results.

4.4.1   DPP Implementation Procedure

To achieve the enriched disparity map, six steps must be executed after performing the DPP method that provides a set of epipolar-rectified images as well as the disparity transferring scales. These six steps, as flowcharted in **Figure 4.8**, include selecting the reference image, generating additional disparity maps, removing disparity outliers and identifying gaps, transferring disparity maps to the reference image (using the disparity transferring scales calculated by DPP), fusing the supplementary disparity maps, and finishing the final enriched disparity map.
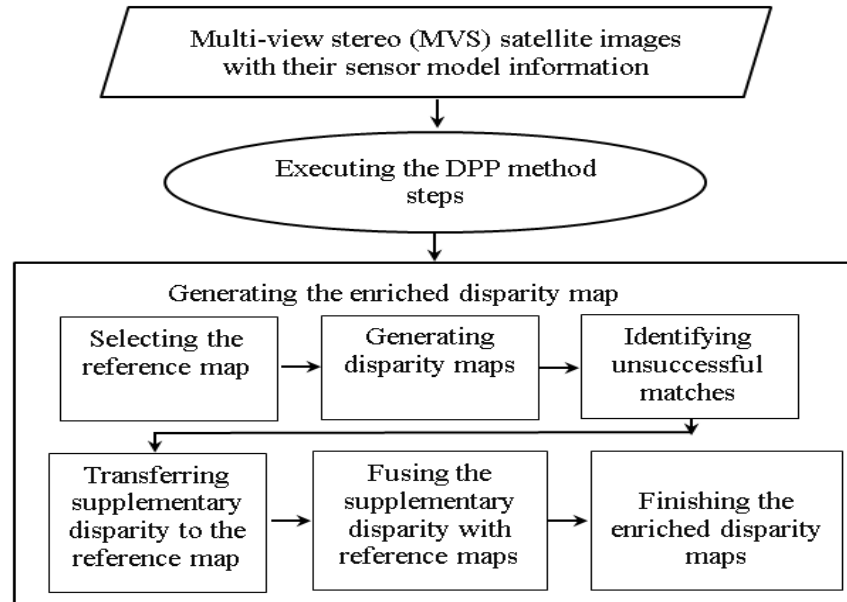
Figure 4.8 DPP-based implementation procedure for generating an enriched disparity map.

4.4.1.1 Selecting the reference image

Different combinations of stereo pairs can be used to generate different disparity maps. Each pair will have different cases of occlusion gaps that need to be filled. Thus, it is recommended to select the image with relatively little occlusions to be the reference image.

4.4.1.2 Generating disparity maps

There are several alternatives for epipolar-based matching algorithms that can be implemented for creating disparity maps from the epipolar images (generated by the DPP method). Alobeid, Jacobsen, and Heipke (2010) concluded that the most suitable matching algorithm for urban areas is semi-global matching (SGM) as introduced in Hirschmüller (2005). Hence, this study adopts this matching algorithm for the generation of disparity maps.

120

4.4.1.3 Identifying unsuccessful matches

This step is required to remove all the unstable (putative) matches and ensure that all the supplementary disparity data are significantly correct. Therefore, a visibility check should be executed to identify the unstable matches and remove them. The Left-Right Checking (LRC) discussed in Egnal and Wildes (2002) is recommended because of its simplicity and accuracy.

4.4.1.4 Transferring the supplementary maps

Transferring the corresponding disparity values to the reference image of the reference stereo domain must be performed before they can be used in gap filling. Hence, all available supplementary disparity maps should be scaled to transfer their disparities to the selected reference disparity map. The transferring scales are calculated by the DPP method based on the object-space offset information in the epipolar direction as described in **Equation (4.5)**.

4.4.1.5 Fusing the supplementary maps

Fusing the transferred supplementary data with the reference disparity map allows filling the gaps and hence enriching the selected reference disparity map. Since the disparity data are measured in the object space and the additional data are scaled to be consistent for the same pixel location, it is preferred for simplicity to be fused in the result level as it is typically followed with object-space elevation data. Several fusion techniques for depth and disparity maps have been proposed. The 3D-median filter fusion technique, described by Zhang et al. (2014), is adopted, with some restrictions applied, in this study for its simple concept, efficiency, and accuracy. It is an extension to the

concept of the 2D-median filter where the third dimension represents the number of the supplementary disparity maps generated.

4.4.1.6 Finishing the enriched disparity map

Even after filling the occlusion gaps and removing outliers after the last step, it is still possible to find very small areas that are without data. If there are no further images to generate extra data, interpolation techniques, or even median filter, can be executed to fill any holes. However, at this time, local neighboring information can be used for gap-filling with minimal risk since the gaps and mismatches are remarkably reduced.

4.4.2 Validation Procedure

The proposed validation procedure in this study includes (V1) confirming the validity of the approximated epipolar lines, epipolarity condition, and the calculated offsets in the epipolar direction, (V2) validating the proportionality of the disparity values, and (V3) quantitatively and qualitatively inspecting the enriched disparity map. Further details regarding these checks (V1, V2 and V3) are provided in the following subsections.

4.4.2.1 Validation of the offset calculation and the epipolarity condition

The confirmation of the calculated offsets is carried out by examining the values of the offset perpendicular to the epipolar direction ($\Delta Y'$) as calculated in **Equation (4.5)**. If these values are all of sub-pixel level, this validates the derived epipolar direction and hence the calculated offsets in the epipolar direction ($\Delta X'$).

Before generating the disparity maps, the epipolarity condition of the rectified images should be validated. This validation can be made by applying an automatic matching technique to find a set of accurate point matches and calculate the differences in the y-disparity (y-parallax) for all point pairs. For this check, the scale-invariant-feature-transform (SIFT) matching technique introduced by Lowe (2004) is selected due to its accuracy and reliability.

4.4.2.2 Validation of disparity inter-proportionality

To confirm and validate the proportionality among the object-space disparity values in the projected images, the epipolar images should meet two conditions: (1) all corresponding image pixels should lie on the same row, and (2) all corresponding image points should lie on one straight line when the epipolar images are arranged vertically by distances proportional to the calculated offsets in the epipolar direction. **Figure 4.9** illustrates this arrangement that satisfies **Equation (4.6)**.
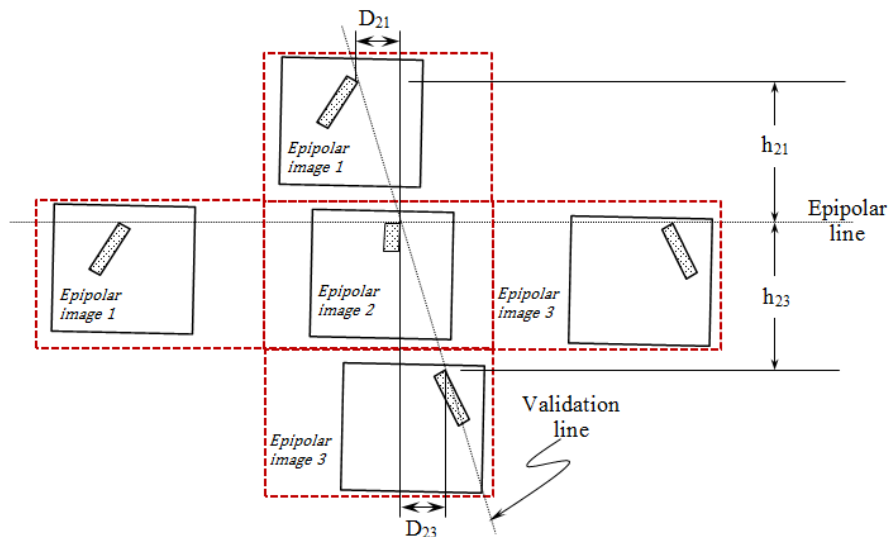
Figure 4.9 Proportionality validation of the achieved object-space disparities.

123

$$\frac{h_{21}}{h_{23}} = \frac{D_{21}}{D_{23}} = \left(\frac{\Delta X'_2 - \Delta X'_1}{\Delta X'_2 - \Delta X'_3}\right) \qquad\qquad (4.6)$$

where h_{21} is the distance between the corresponding epipolar lines in the epipolar image (2) and image (1). D_{21} is the object-space disparity between the corresponding matching points in the epipolar image (2) and the epipolar image (1). The same analogy applies for h_{23} and D_{23}. $\Delta X'_1, \Delta X'_2$, and $\Delta X'_3$ are the calculated offsets with respect to the epipolar direction of the images (1, 2, and 3) respectively.

4.4.2.3 Disparity map validation

The disparity map should be validated both quantitatively and qualitatively. Quantitative evaluation of the calculated disparity values can be conducted by deriving their corresponding elevations and comparing these elevations against accurate reference elevation data. Unlike LiDAR-derived elevations, photogrammetrically-derived elevations are feasible and available for most places. If multiple stereo images, their sensor model information, and a set of ground control points (GCPs) are available for the study area, an accurate digital surface model (DSM) for the study area can be generated photogrammetrically by the traditional bundle adjustment as available in the commercial software packages and then used as a reference elevation data for accuracy evaluation.

Qualitative assessment based on visual inspection is proposed to identify the homogeneity of the values related to the same object. For further examination, the generated disparity map can be rendered using the relevant reference imagery for 3D representation since it is perfectly co-registered with the selected reference epipolar

image. The quality of the achieved result can then be directly and easily validated, and any inconsistent values can be straightforwardly identified.

4.5   Data, Results, and Discussion

4.5.1   Dataset and Study Area

The optical data used in this research are a subset of tri-stereo VHR images acquired by linear sensor with push-broom scanning mode. These stereo images were taken by the Pleiades-1A satellite over the urban area of Melbourne, Australia (2012). The two opposite acquisition angles of the off-nadir images are approximately +15 and -15 degrees respectively. The ground resolutions are 0.5 m/pixel for the panchromatic and 2 m/pixel for the four multispectral bands.

This test area was specifically selected to represent a dense urban environment with a variety of building shapes, sizes, and heights. The study area spans about 1.5km by 0.8km and contains 170 buildings, 15 of which are high-rise buildings.

4.5.2   Results and Validation of DPP Method

The selected horizontal projection plane for the study area was at 35m above the WGS84 ellipsoid which represents the average elevation of the terrain. The second parallel projection plane was selected randomly at the datum (0m level). The offsets $(\Delta X, \Delta Y)$ calculated by **Equation (4.3)** were rotated by angle θ into the corresponding offsets $(\Delta X', \Delta Y')$ relative to the epipolar direction using **Equation (4.4)**. The same value for the epipolar angle θ was derived and used in the three employed

satellite stereo images which is expected for the inline tri-stereo satellite images. **Table 4.1** provides the minimum and the maximum values for all of the computed offsets.

Generally for a large dataset, the offsets $(\Delta X')$ should be used as calculated for each single ground pixel. However, the study area is of a small size. This situation led to very small variations in the calculated offsets as shown in the $\Delta X'$ column of **Table 4.1**. For this reason, this study used the average values for the offsets to simplify the calculations. Accordingly, the value for $\Delta X'_1$ for all pixels in the first image (backward off-nadir image) was 10.371 pixels, $\Delta X'_5$ for the second image (nadir image) was -0.007 pixels, and $\Delta X'_3$ for the third image (forward off-nadir image) was -10.118 pixels.

As shown in **Table 4.1**, the achieved values for the computed offsets $(\Delta Y')$ were all in the sub-pixel range (≤ 1.0 pixel). This confirms the validity (V1) of the derived epipolar direction and the calculated offsets in the epipolar direction$(\Delta X')$. At the same time, the variations of the epipolar direction offsets $(\Delta X')$ were almost zero, which agrees with the expectations regarding the size of the test data. Since the disparity transferring scales are calculated for each pixel, the technique is applicable even for large areas as long as the epipolarity condition is preserved. This epipolarity condition has been proved by Wang, et al. (2011) for a variety of image data including long and full size scenes over hilly and mountainous areas.

To confirm the epipolarity condition of the epipolar rectified images, the SIFT matching technique was executed to generate a set of 30 matching points of good distribution. Then, the differences in the matching location were calculated between the point pairs to assess the quality of the epipolarity condition. The absolute average of these differences was found to be 0.32 pixels with a standard deviation of 0.22 pixels.

These values validate the approximation of the epipolar curves and indicate the success of the epipolar rectification for the employed MVS-VHR satellite images.

Table 4.1 The ranges for the computed offsets in the original planimetric directions $(\Delta X, \Delta Y)$ and with respect to the epipolar direction $(\Delta X', \Delta Y')$.

Image Number	ΔX [Min, Max]	ΔY [Min, Max]	$\Delta X'$ [Min, Max]	$\Delta Y'$ [Min, Max]
Backward off-nadir (I1)	[2.820, 2.912]	[9.985, 9.994]	[10.359, 10.383]	[-0.715, -0.625]
Nadir Image (I5)	[0.659, 0.747]	[-0.162, -0.159]	[-0.017, 0.002]	[-0.800, -0.678]
Forward off-nadir (I3)	[-1.466, -1.374]	[-10.050, -10.042]	[-10.129, -10.106]	[-0.765, -0.709]

*All values are in pixels. The ground pixel size is 1 m^2

At this stage, one can validate the disparity inter-proportionality (V2). For this purpose, the ratio of the distances between corresponding epipolar lines needs to be calculated using **Equation (4.6)**. The ratio of $(S = h_{51}/h_{53})$ was equal to -1.026. The negative sign of this ratio indicates that the off-nadir images are of opposite viewing directions. The epipolar images are arranged based on the calculated ratio. **Figure 4.10** shows that all of the corresponding pixels lie on the same row horizontally satisfying the epipolarity condition. In this figure, for example, while the outer pixels' coordinates– measured from the upper-left origin–of the upper epipolar line are (167.5, 336.5) and (1201.5, 336.5), the outer pixels of the lower epipolar line are (1506.4, 427.5) and (475.5, 427.6). At the same time, the corresponding pixels lie on the same inclined straight line when the epipolar images are arraigned vertically. Quantitatively, while the matching pixel coordinates on the left validation line in **Figure 4.10** are (728.4, 69.5), (684.38, 336.5), and (640.5, 595.9); the pixels on the right validation line are (1036.4, 160.5), (990.5, 427.6), and (946.4, 686.6). This situation –in both horizontal and vertical directions–validates the equations derived based on the calculated offsets (i.e., **Equation**

127

(4.2) and **Equation (4.6)**). Moreover, it validates the inter-proportionality (scale-relation) of the disparity values (V2) for all corresponding pixels. Thus, the disparity can now be transferred from an image domain to another one for gap filling.
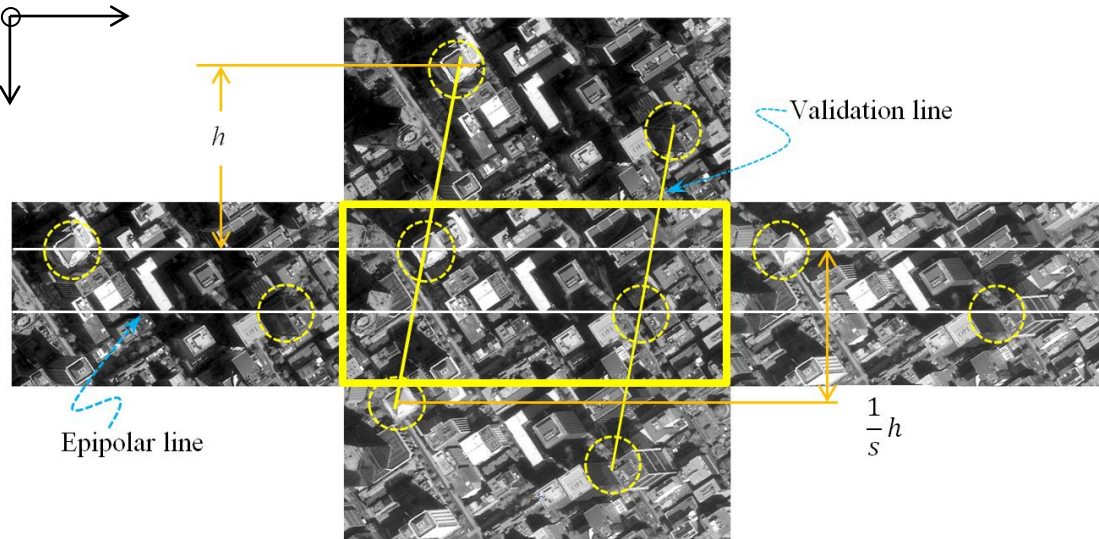
Figure 4.10 Validation of the disparity values proportionality; the inclined vertical lines indicate the proportionality of the disparity values since all corresponding ground pixels lie on the same straight line. The horizontal lines represent the epipolar lines.

4.5.3   Generating and Validating the Enriched Disparity Map

This section provides some details on the execution and the results after implementing the six steps of the DPP-based implementation procedure. To achieve the enriched disparity map, the nadir image (I5) was selected to be the reference image. This is because most of the occluded areas in the forward off-nadir image (I3) were imaged in the backward one (I1) and vice versa. The SGM matching algorithm was executed for the nadir image against the backward and forward off-nadir stereo images. The stereo pair of the nadir and backward images (i.e., images I5 and I1) was selected to be the reference stereo domain.

The generated disparity map (D_{51}) for the nadir-forward image domain (I5-I1) is shown in **Figure 4.11-(a)** and the disparity map (D_{53}) for the nadir-backward image domain (I5-I3) is shown in **Figure 4.11-(b)**. Next, the Left-Right visibility check was executed on the two stereo pairs to identify disparity gaps and outliers. Pixels in the generated disparity maps with such deficiencies are represented in white in **Figure 4.11-(c and d)** for the reference and the supplementary stereo pair respectively. The effect of the occlusion is clear in these two figures since almost one-fifth of the generated maps are occluded. Quantitatively, while the gaps in **Fig. 11(c)** represent 21.1%, the gaps in **Fig. 11(d)** represent 19.7%.

Figure 4.11-(e) shows the pixels that can be filled from the supplementary domain (I5-I3). These pixels represent 15.2%. In contrast, **Figure 4.11-(f)** shows the gaps that must be interpolated since no data were available in the supplementary domain. Despite the shortage in the supplementary data, the gaps in **Fig. 11(f)** represent only 4.9% of the reference image. This indicates a reduction of 77% in the interpolation risks if executed in such urban area. After that, the disparity data were then scale-transferred from the supplementary domain (D_{53}) to the reference domain (D_{51}) based on the offset values calculated in **Table 4.1** and **Equation (4.4)**. The result of this step is illustrated in **Figure 4.11-(g)**.

Prior to fusing the two disparity maps and adopting specific values, a consistency check was applied to compare the disparity values in the reference map against the corresponding ones in the supplementary map. **Figure 4.11-(h)** shows those pixels that had disparity values in both domains and within 2 pixels of each other. In this case, the values of the reference disparity map were retained. On the other hand, the median (based

on a 3D-median filter) was adopted for pixels with disparity values that differed by more than the selected threshold. These pixels, as shown in **Figure 4.11-(i),** represent only 3.2% of the reference image. This very small value indicates the success on the Left-Right checking technique used to exclude the unstable (putative) matches.

Two other possible cases still exist. One case is for pixels with no-data in the reference disparity map but with values in the supplementary disparity map, for which the values in the supplementary disparity map were directly accepted. The final case is for pixels with no data in both domains. In such a case, the gaps were interpolated from the surrounding values in both maps using a 3D-median filter. Hence, this filter was applied only on the pixels shown in **Figure 4.11-(f,** and **i)** based on the data available in **Figure 4.11-(a,** and **g)**. The final disparity map achieved is shown in **Figure 11-(j)**. Clearly, this map is noise free and has no gaps and, as such, satisfies the established disparity map visual validation.

To quantitatively evaluate the quality of the achieved disparity map in **Fig. 11(j)**, the DSM elevations corresponding to this map were triangulated. After that, the original full-scene stereo pairs, their sensor model information, and a set of 7 GCPs were all employed to generate photogrammetrically an accurate DSM for the study area using a specialized commercial software package (PCI Geomatica). Based on 21 independent check points (10-15 cm accuracy), this DSM showed a root-mean-square-error (RMSE) of 0.72m and 1.36m in both horizontal and the vertical directions respectively. Hence, the achieved DSM was adapted as a reference elevation data. After that, the average elevations of 29 building rooftops were selected manually within the test area in both the derived DSM

elevations from the enriched disparity map and the generated reference DSM. The resulted RMSE was found to be 1.02 meter.
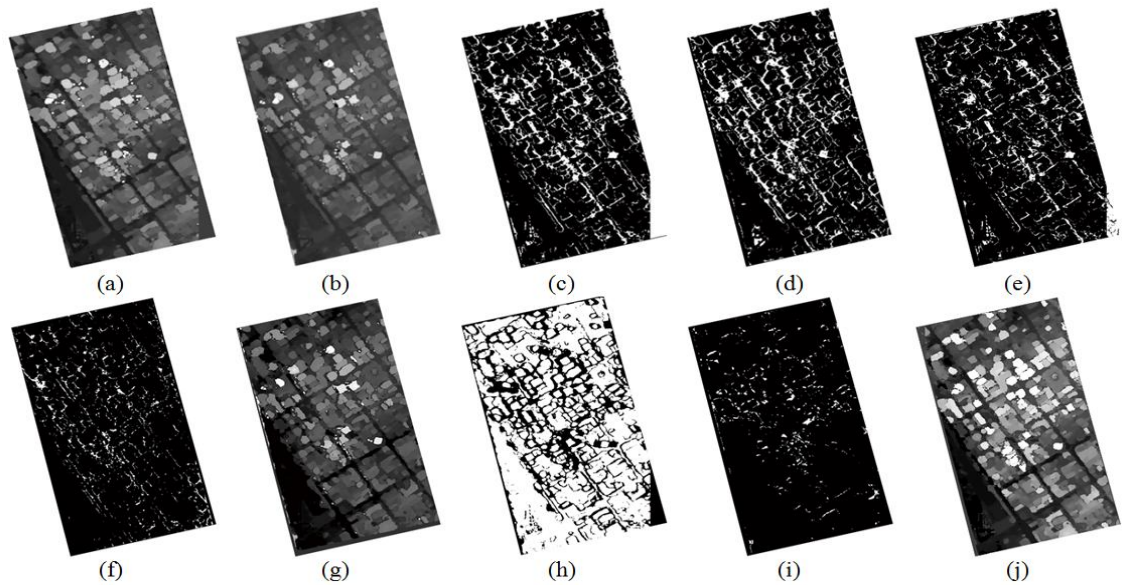
Figure 4.11 Steps and intermediate results in the DPP-based generation of enriched disparity map; (a) disparity map of the reference domain I5-I1, (b) disparity map of the supplementary domain I5-I3, (c) and (d) gaps (in white) in the first and second disparity maps, respectively, (e) pixels that can be filled from the supplementary disparity data (in white), (f) pixels that cannot be filled from the supplementary disparity data (in white), (g) result of transferring the supplementary domain into the reference domain, (h) transferred pixels with correct disparity values after a consistency check (in white), (i) pixels failed the consistency check (in white), (j) the final enriched disparity map.

Visual inspection was adopted for the qualitative assessment of the achieved disparity map. **Figure 4.12** shows different isometric views for the generated 3D rendered representation of the study area. **Figure 4.12-(b1, b2, c, d,** and **e)** clearly shows the high quality of these 3D representations despite the structural complexity of the study area. The figure also illustrates the effect of the achieved enhancement in comparison to the original un-enhanced disparity data that include the gaps and outliers shown in **Figure 4.12-(a1,** and **a2)**.
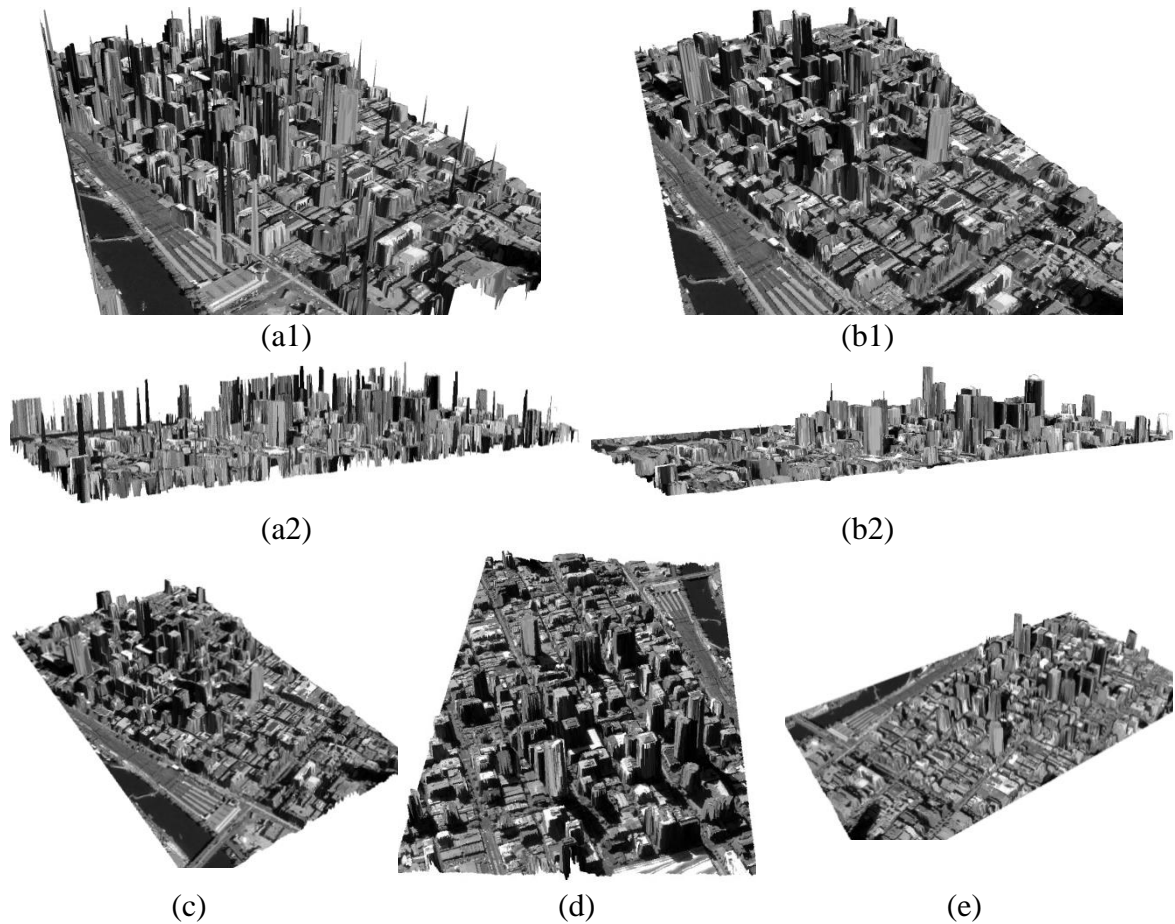
Figure 4.12 Different isometric views based on disparity maps before and after enrichment. The sub-figures (a1) and (a2) are rendered based on un-enhanced disparity map. The sub-figures (b1) and (b2) show the result after DPP-based production of the enriched disparity map. The sub-figures (c), (d) and (e) are further isometric views reflecting the achieved quality of the enriched disparity map.

4.6   Conclusions

Motivated by the advantages of using disparity data instead of elevation data, this study endeavored to address the problem of filling the gaps in disparity maps generated from high resolution satellite images. The study proposes the adoption of a disparity map that is enriched by supplementary disparity data generated from multi-view stereo satellite images. The key concept for the proposed solution is to project inline MVS satellite images onto a relevant object space horizontal plane to ensure consistency

among the disparity values in the object space. For that purpose, a method based on Double Projection Planes was developed. The novelty of this method is in exploiting the sensor model information for extracting offset information based on a two-plane projection process in order to determine the disparity transferring scales among different image stereo domains.

After executing the designed implementation procedure, the achieved result based on the developed DPP method was found highly successful. The proportionality of the disparity values was validated successfully using a vertical arrangement of the constructed epipolar images. This arrangement is based on calculating the object-space offsets and transferring scales. By visual inspection, the enriched disparity map produced 2D and 3D representations of high quality with homogeneous and flat building roofs. Quantitatively, the elevations derived based on this enriched map were found of almost pixel-level (metre) accuracy when evaluated against DSM elevations derived photogrammetrically using a specialized commercial software package

One may confidently conclude that the developed DPP-based method is effective and capable of providing gap-and-outliers-free disparity maps from inline MVS satellite images. Such disparity maps can be used to replace or reduce the need for DEMs in the processing of satellite images for extracting information.

Future research based on this study will test the applicability of the DPP-based enriched disparity maps in a building detection process using off-nadir VHR satellite imagery acquired over dense urban areas.

ACKNOWLEDGEMENTS

REFERENCES

Alobeid, A., Jacobsen, K., & Heipke, C. (2010). Comparison of matching algorithms for DSM generation in urban areas from Ikonos imagery. *Photogrammetric Engineering and Remote Sensing, 76*(9), 1041-1050.

Cho, S., Sun, I., Ha, J., & Jeong, H. (2013). Occlusion removal and filling in disparity map for real time multiple view synthesis. *Journal of Communications and Information Sciences, 3*(3), 425-432.

Egnal, G., & Wildes, R. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24*(8), 1127-1133.

Gehrke, S., Downey, M., Uebbing, R., Welter, J., & LaRocque, W. (2012). A multi-sensor approach to semi-global matching. *International Archives of Photogrammtry, Remote Sensing, and Spatial Information Sciences*, *39*(B3), 17-22. doi: 10.5194/isprsarchives-XXXIX-B3-17-2012

Goesele, M., Curless, B., & Seitz, S. (2006). Multi-view stereo revisited. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York City, NY, USA

Grodecki, J., Dial, G., & Lutes, J. (2004). Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. *Proceedings of the ASPRS Annual Conference*, Denver, Colorado, USA.

Habib, A., Morgan, M., Jeong, S., & Kim, K. (2005). Analysis of epipolar geometry in linear array scanner scenes. *The Photogrammetric Record, 20*(109), 27-47. doi: 10.1111/j.1477-9730.2005.00303.x

Hirschmüller, H. (2005). Accurate and efficient stereo processing by Semi-Global Matching and Mutual Information. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA.

Hirschmüller, H., (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2): 328–341.

Hosni, A., Bleyer, M., Gelautz, M., & Rhemann, C. (2009). Local stereo matching using geodesic support weights. *Proceedings of the 16th IEEE International Conference on the Image Processing (ICIP 09)*, Cairo, Egypt.

Huq, S., Koschan, A., & Abidi, M. (2013). Occlusion filling in stereo: Theory and experiments. *Computer Vision and Image Understanding, 117*(6), 688-704.

Krauß, T., & Reinartz, P. (2010a). Enhancement of dense urban digital surface models from VHR optical satellite stereo data by pre-segmentation and object detection. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(WG1/4): 6 pages, Calgary, Canada.

Krauß, Thomas, & Reinartz, Peter. (2010b). Urban object detection using a fusion approach of dense urban digital surface models and VHR optical satellite stereo data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(WG 1/4): 6 pages, Istanbul, Turkey.

Krauß, Thomas, Sirmacek, Beril, Arefi, Hossein, & Reinartz, Peter. (2012). Fusing stereo and multispectral data from WorldView-2 for urban modeling. *The International Society for Optics and Photonics*, (pp. 83901X–83901X), Baltimore, Maryland, USA.

Lowe, G., (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Merrell, P., Akbarzadeh, A., Wang, L., Mordohai, P., Frahm, J., Yang, R., . . . Pollefeys, M. (2007). Real-time visibility-based fusion of depth maps. *Proceedings of the IEEE International Conference on Computer Vision (ICCV 07)*. Rio de Janeiro, Brazil.

Min, D., & Sohn, K. (2008). Cost aggregation and occlusion handling with WLS in stereo matching. *IEEE Transactions on Image Processing, 17*(8), 1431-1442.

Morgan, M., Kim, K., Jeong, S., & Habib, A. (2006). Epipolar resampling of space-borne linear array scanner scenes using parallel projection. *Photogrammetric Engineering and Remote Sensing, 72*(11), 1255-1263.

Oh, J., & Kuo, C-C. (2010). Robust stereo matching with improved graph and surface models and occlusion handling. *Journal of Visual Communication and Image Representation,* 21(5), 404-415.

Strecha, C., Fransens, R., & Van Gool, L. (2006). Combined depth and outlier estimation in multi-view stereo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 06)*. (pp. 2394–2401), New York City, USA.

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016a). Enhanced Disparity Maps from Multi-View Satellite Images. *The 2016 IEEE Geoscience and Remote Sensing Symposium (IGARSS 16)*, Beijing, China.

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016b). Registration-based Mapping of Aboveground Disparities (RMAD) for Building Detection in Off-nadir VHR

Stereo Satellite Imagery *Photogrammetric Engineering and Remote Sensing, 82*(7), 535-546.

Szeliski, R. (1999). A multi-view approach to motion and stereo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 99)*, (pp. 157–163), Fort Collins, USA.

Tao, C., & Hu, Y. (2001). A comprehensive study of the rational function model for photogrammetric processing. *Photogrammetric Engineering and Remote Sensing, 67*(12), 1347-1358.

Tao, C., & Hu, Y. (2002). 3D reconstruction methods based on the rational function model. *Photogrammetric Engineering and Remote Sensing, 68*(7), 705-714.

Unger, C., Wahl, E., Sturm, P., & Ilic, S. (2012) Stereo fusion from multiple viewpoints. *Lecture Notes in Computer Science (LNCS),* 7476, 468-477.

Wang, M., Hu, F., & Li, J. (2011). Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 347-355. doi: 10.1016/j.isprsjprs.2011.01.002

Zhang, J., Cao, Y., Zheng, Z., Chen, C., & Wang, Z. (2014). A new closed loop method of super-resolution for multi-view images. *Machine Vision and Applications, 25*(7), 1685-1695.

Chapter 5: DISPARITY-BASED GENERATION OF LINE-OF-SIGHT DSM FOR IMAGE-ELEVATION CO-REGISTRATION TO SUPPORT BUILDING DETECTION IN OFF-NADIR VHR SATELLITE IMAGERY[1]

Abstract

The integration of optical and elevation data is of great importance for 3D-assisted mapping applications. Very high resolution (VHR) satellite images provide ideal geo-data for mapping building information. Since buildings are inherently elevated objects, these images need to be co-registered with their elevation data for reliable building detection results. However, accurate co-registration is extremely difficult for off-nadir VHR images acquired over dense urban areas. Therefore, this research proposes a Disparity-based Elevation Co-Registration (DECR) method for generating a Line-of-Sight Digital Surface Model (LoS-DSM) to efficiently achieve optical-elevation data co-registration with pixel-level accuracy. Relative to the traditional photogrammetric approach, the RMSE value of the derived elevations is found to be less than 2 pixels. The applicability of the DECR method is demonstrated through elevation-based building detection (EBD) in a challenging dense urban area. The quality of the detection result is found to be more than 90%. Additionally, the detected objects were geo-referenced

---

successfully to their correct ground locations to allow direct integration with other maps. In comparison to the original LoS-DSM development algorithm, the DECR algorithm is more efficient by reducing the calculation steps, preserving the co-registration accuracy, and minimizing the need for elevation normalization in dense urban areas.

5.1 Introduction

Since buildings are the predominant object class in urban areas, building information is crucial for city planning and management applications. The commonly used data for extracting this type of information are the very high resolution (VHR) satellite images. Since buildings are inherently elevated objects, the integration of optical images and elevation data is of great importance for building detection applications. Accordingly, elevation-based building detection in VHR satellite images has become an active area of research in the remote sensing community.

VHR optical satellite images have been acquired since 1999 by several earth observation systems. These images are characterized by broad coverage, sub-meter ground resolution, relatively low cost, and rich information content that is necessary for mapping complex urban environments. Additionally, due to the agility of the currently available satellite sensors, VHR images are usually acquired off-nadir with across-track and/or along-track angles to provide fast and various acquisition modes including mainly the stereo mode.

Stereo images are among the highly demanded types of images. This is because they allow generating stereo-based information (elevations and disparity information) which facilitates identifying elevated urban features. For urban mapping applications, elevation

information plays a fundamental role in elevation-based building detection in VHR images. This is because the information about building heights improves the detection results by allowing accurate detection and reliable distinguishing of the building-roofs from even spectrally and/or spatially similar objects in urban environments such as parking lots.

Digital elevation models (DEMs) are commonly used to store elevation information in different forms. These forms include digital terrain models (DTMs) describing the terrain elevations, digital surface models (DSMs) describing the visible surface elevations (both terrain and off-terrain elevations), and the normalized digital surface model (nDSM) describing the aboveground elevations. However, these DEM forms represent only the geometric information and lack the rich information content of the optical VHR images (specifically radiometric, texture, spectral, and contextual information). Therefore, when DEM data are generated, they need to be integrated with their relevant optical data to achieve a more complete dataset for 3D-assisted information extraction applications.

The photogrammetric-based generation of elevation data from stereo VHR images is achievable at a cheaper rate than other remote sensing sources including LiDAR (Light Detection and Ranging) technology. The generated stereo-based elevation data represent the height information of the tops of surfaces such as buildings and trees; hence, they result in DSMs. Currently, VHR satellite sensors can produce multi-view inline stereo (MVIS) images for the same region using the off-nadir acquisition capability. These MVIS-VHR satellite images provide comprehensive coverage, minimize the occlusions in the complex environments, and allow dense and accurate DSMs to be generated.

139

The integration of the optical and elevation datasets leads to higher quality surface information since these two data types have complementary properties. Therefore, co-registering these two datasets is a valid integration approach to make one data source that compensates for the weaknesses of both. However, several problems are introduced when such multi-data sources are integrated, for instance, the problematic misregistration of DSMs when they are co-registered to off-nadir VHR images. This problem is due to the severe perspective and relief displacement effects caused by the off-nadir acquisition of the VHR satellite images when captured over dense urban areas (Salehi, Zhang, & Zhong, 2011; Salehi, Zhang, Zhong, & Dey, 2012).

Bearing in mind this problem of optical-elevation data co-registration, our study reviewed the relevant literature. The line-of-sight DSM (LoS-DSM) solution introduced by Suliman and Zhang (2015) provides accurate optical-elevation co-registration even for the off-nadir VHR satellite images more efficiently than the other available techniques. In this research, the aim is to improve the efficiency of this solution further by minimizing the computation steps, keeping the pixel-level co-registration accuracy, and reducing the need for elevation data normalization in dense urban areas.

Stereo images allow generating two types of co-related surface information: surface disparity maps (SDMs) and surface elevation models (DSMs). While the SDMs represent the relative measurements in the image spaces of one stereo pair, DSMs represent the absolute derived elevations in the object space. Disparity maps are computationally cost-effective and perfectly co-registered to the reference image. Hence, the use of such maps should satisfy the aspects of the research aim if the disparity data are elevation scale-related and of minimized terrain effects. Thus, in order to achieve the proposed three-

aspect aim, we propose extending the double projection planes (DPP) method, introduced by Suliman and Zhang (2016), for deriving LoS-DSM elevations from an enriched disparity map that is directly co-registered to the reference image with pixel-level accuracy. Furthermore, this map should have minimized terrain-level disparities.

The proposed extension of the DPP method leads to the Disparity-based Elevation Co-Registration (DECR) method for efficiently generating LoS-DSM that is co-registered with pixel-level accuracy to the reference image. The ultimate goal of this research is to demonstrate the applicability of the DPP-based improved solution (DECR method) through an elevation-based building detection in off-nadir VHR satellite images acquired over challenging urban areas. The goal includes also geo-referencing the mapped rooftops in an off-nadir image for direct integration in geographic information systems (GIS).

The remainder of this paper is organized as follows: **Section 5.2** reviews the previous work and identifies the research problem; **Section 5.3** describes the proposed DECR method; **Section 5.4** describes the DECR method validation procedure; **Section 5.5** demonstrates the applicability of the improved LoS-DSM solution using the DECR method in an elevation-based building detection; and **Section 5.6** describes the data used and discusses the achieved experimental results. Finally, the conclusions are drawn in **Section 5.7**.

5.2   Previous Work

As can be seen in the relevant literature, image-to-image registration methods are used widely to address the problem of accurate optical-elevation data co-registration.

These methods, by considering the elevation data as 2.5D maps, aim to determine the point-by-point correspondence between two images of a scene by implementing four fundamental steps: feature extraction, feature matching, transformation model estimation, and resampling (Goshtasby, 2012). Since image and elevation data have different characteristics and appearances, the main challenge of these methods is detecting common objects in the images being co-registered (Mishra & Zhang, 2012). Accordingly, co-registration methods, as reviewed extensively by Zitová and Flusser (2003), are categorized generally as either feature-based or intensity-based methods. While feature-based methods establish the registration between the two datasets via matching common features, intensity-based methods utilize intensity values within the datasets to find similarity. The resampling process in most registration methods is applied to the elevation data, instead of the image data, in order to avoid producing radiometric distortions that may affect the subsequent processes.

On the one hand, the common features used in feature-based methods include point, line, or planar/area features (Xiong & Zhang, 2010). Point features extracted by a corner point detector were used by Wong and Orchard (2008) to register an optical image to its corresponding LiDAR data. Line features were used by Habib, Ghanma, Morgan, and Al-Ruzouq (2005) and Avbelj, Iwaszczuk, Müller, Reinartz, and Stilla (2015) to co-register both image and elevation data. For extracting building roofs using LiDAR data and multispectral imagery, Kwak, Kim, Yu, and Lee (2006) extracted the roof planes from both optical and LiDAR data and then co-related their centroids. On the other hand, intensity-based registration methods evaluate statistical dependency between the image and elevation datasets in order to derive similarity measures. These methods avoid the

complex processing of feature extraction and matching, and hence provide better results. Mutual information, as the most successful example, has been proven to be effective for the optical and elevation data registration as demonstrated by Parmehr, Fraser, Zhang, and Leach (2014).

All of the image-to-image registration methods involve expensive computations during the four fundamental steps listed earlier. Moreover, in dense urban areas, most of these registration methods encounter major difficulties when off-nadir VHR satellite images are employed due to the severe building leaning caused by the off-nadir perspective acquisition of these images. In contrast to elevation models that have orthographic projection, the building leaning in the perspective mages creates building facades that represent objects not included in the orthographic elevation models. Hence, unless the geometry information of the image acquisition is incorporated in the co-registration process, this projection difference in the two datasets poses a standing problem.

True orthorectification can provide accurate optical-elevation co-registration by generating orthoimages corrected for relief displacements and perspective effects. Many studies have benefited from the direct co-registration between the orthoimages and their corresponding elevation data for information extraction (e.g., Awrangjeb, Ravanbakhsh, & Fraser, 2010; Awrangjeb, Zhang, & Fraser, 2013). True orthorectification procedures, unlike most image-to-image registration methods, bypass the steps of finding common features and estimating transformation model parameters, required in conventional registration methods, by utilizing the image acquisition information. Guoqing, Weirong, Kelmelis, and Deyan (2005) provided a comprehensive study on the urban true

orthorectification process. In this process, elevation data are used and the sensor model information of the employed perspective images are incorporated to eliminate the effects of the off-nadir perspective image acquisition and hence perfectly align image to elevation data (Habib, Kim, & Kim, 2007). However, true orthorectification requires digital building models which are not available for most places. Additionally, it requires sophisticated algorithms for occlusion detection and compensation to fill the occlusion gaps. Furthermore, radiometric enhancement techniques are needed to correct and balance the resampled brightness values and produce seamless mosaics of true orthoimages. Hence, true orthoimages are expensive, time consuming, and difficult to achieve.

It has been concluded that all of the reviewed methods are inefficient–in terms of implementation cost and result quality–when off-nadir images over dense urban environments are employed. In contrast, a line-of-sight DSM (LoS-DSM) solution has been introduced recently by Suliman and Zhang (2015) to address the problem of efficient image-elevation co-registration based on incorporating the sensor model information of the images used. Thus, in this study, we are reviewing the LoS-DSM solution to identify aspects for improvement since it is a simple, promising, and recent technique for the problem posed in our research.

The LoS-DSM solution was developed to achieve accurate co-registration and preserve the original image information. It was proposed to mitigate the limitations of the other currently available methods. The solution algorithm is straightforward, efficient, and effective even when off-nadir images over dense urban environments are employed. The solution effectiveness has been demonstrated through elevation-based building

detection in urban off-nadir VHR imagery. Nevertheless, we argue that the algorithm efficiency can be improved in terms of two aspects: minimizing the required computation steps and reducing the need for elevation normalization.

Regarding the computation steps, this algorithm incorporates the sensor model information of the VHR images in two major steps: triangulating the elevation data photogrammetrically (space intersection) and then re-projecting these elevations back to the image space (co-registration) to achieve accurate registration and original image data preservation. Hence, the algorithm has redundant calculation steps that would be efficient if they are bypassed and, at the same time, the pixel-level co-registration accuracy is maintained. These redundant steps include the space intersection, bundle adjustment, and 3D coordinates' calculation as well as the back re-projection of the triangulated DSM elevations.

Additionally, since the elevation data are transferred to the original image space, the terrain relief distortions in the off-nadir images are not eliminated or minimized. Hence, a normalization process for the co-registered DSM elevations, to generate the nDSM, cannot be ignored for dense urban areas even of moderate terrain variation which is usually the case based on the geometric road design standards in urban areas. This process requires two subsequent steps: extracting the DTM from DSM data and then subtracting the resulting DTM from its corresponding DSM ($nSDM = DSM - DTM$). The extraction and subtraction steps usually introduce elevation errors that negatively affect the results of the 3D-based mapping applications. In the case of employing off-nadir VHR images acquired over dense urban areas, these errors become extremely

serious due to the prominent existence of the building lean (Suliman, Zhang, & Al-Tahir, 2016).

Although the LoS-DSM solution was demonstrated to be successful even for off-nadir images, the original algorithm still has room for further improvement. The identified areas of improvement lie in the steps of elevation generation, co-registration, and normalization. Hence, the aspects of the research problem include (1) efficient generation of elevation data from MVIS-VHR satellite images, (2) direct optical-elevation data co-registration of pixel-level accuracy with one of the employed images, and (3) minimized terrain relief variation of the co-registered elevations to reduce the need for elevation normalization.

The photogrammetric approaches, as executed in the original LoS-DSM algorithm, involve computationally expensive steps such as space intersection and bundle adjustment since the image-space SDMs and object-space DSMs are typically not scale co-related. Hence, if a scale relationship is achieved between SDMs and DSMs, an efficient generation and accurate co-registration of LoS-DSM will be accomplished since the SDMs, as per definition, are co-registered exactly to the reference image selected. Moreover, if the SDMs are measured from epipolar stereo images co-registered based on terrain-level features as demonstrated by Suliman et al. (2016), the mapped SDMs will have minimal terrain effects. As a result, the derived DSM elevations from these SDMs should have minimal terrain effects that can be ignored to bypass the typical steps of DSM normalization in elevation-based building detection in dense urban areas.

Thus, we propose in this research to improve the three identified aspects of the research problem by deriving elevation data (DSM) from enriched disparity maps of

minimized terrain disparities in the relevant object space based on a scale relationship. This allows direct elevation data generation and accurate co-registration even with off-nadir VHR images. Since the foundation for this solution is provided by Suliman and Zhang (2016), extending their DDP method to derive elevations should provide an improved solution compared to the original LoS-DSM algorithm.

5.3   Disparity-based Elevation Co-registration (DECR) Method

The DECR method aims to generate efficiently a LoS-DSM co-registered to the relevant epipolar image data with pixel-level accuracy. The concept is to achieve a dense disparity map in the object space that has a scale relationship to the corresponding elevations. Since the generated disparity maps fit exactly their reference images as per definition, converting the co-registered disparity values to their corresponding elevation will result in perfectly co-registered elevations to the same reference images. Hence, elevation values are generated and co-registered based on the proportionality with their corresponding disparity values.

The DECR method represents an extension of the DPP method described in Suliman and Zhang (2016). The DPP method utilizes double projection planes in the object space for co-relating the corresponding disparities from different epipolar stereo pairs to generate an enriched disparity map. Thus, this co-relation can be extended to derive elevation values from their corresponding disparities. The DECR method, as flowcharted in **Figure 5.1**, comprises four steps: executing the DPP method to achieve supplementary disparity maps, generating an enriched SDM from the supplementary data, calculating the

conversion scales, and deriving the co-registered elevation model (i.e., disparity-based LoS-DSM). Further details are provided in the following subsections.
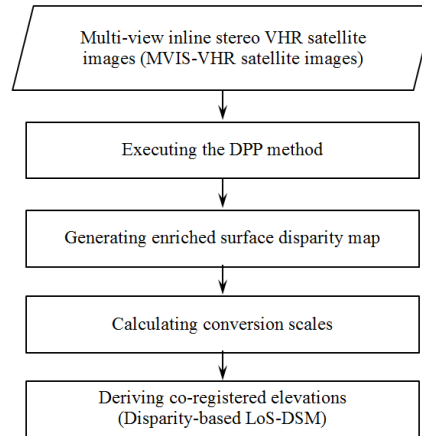
Figure 5.1 The flowchart of the disparity-based elevation co-registration (DECR) method for generating disparity-based LoS-DSM from MVIS-VHR satellite images.

5.3.1 Executing the DPP Method

As a brief description, the double projection plane (DPP) method exploits the sensor model information (ground-to-image transformation and its opposite form) of the employed VHR satellite images for projecting these images onto two horizontal and parallel object-space planes. The first plane is at the average terrain level of the imaged area (Z_{avg}) in order to minimize the terrain relief displacements and to approximately straighten the epipolar lines of the push-broom linear sensor that is the typical type for acquiring VHR satellite images. Additionally, this projection allows object-space disparities to be scale-related among different stereo pairs. The second plane is arbitrarily chosen at the datum (zero level) of the imaged area in order to build geometrical relationships based on computed offsets for deriving the transferring scales for the disparity values among different stereo domains. Readers are encouraged to consult the detailed description provided in Suliman and Zhang (2016) of how to derive the

148

transferring scales. The product of executing this method is epipolar images along with shift information for calculating the transferring scales among different disparity maps.

5.3.2   Generating Enriched Surface Disparity Map

By now, the VHR satellite images are projected into the two plane projection in the object space and the offset information for each projected image is computed. After that, six steps are then executed as implemented in Suliman and Zhang (2016). The product of implementing these steps is a significantly gap- and outlier-free disparity map which has minimized terrain relief effect.

These steps are started by selecting the reference image, which is recommended to have minimal occlusion in the image dataset employed. The second step is to execute an effective matching technique for generating several disparity maps from different combinations of the epipolar pairs of the projected images (the reference image should be kept the same for all combination to have different disparity maps with the same reference frame). Then, the unsuccessful and putative matches resulting from the matching process need to be excluded. Hence, a visibility checking technique like those reviewed in Egnal and Wildes (2002) can be implemented (e.g., the Left-Right Checking technique). Afterward, the disparity transferring scales are derived from the computed shift information as described in the DPP method. These scales are used to transfer the extra disparity maps from their image domains to the domain of the selected reference image. Hence, the fifth step is to fuse the resulting supplementary data for enriching the disparity map of the selected reference image. The 3D-median filter fusion technique, as

introduced by Zhang, Cao, Zheng, Chen, and Wang (2014) is recommended. Finally, the achieved disparity map after the fusion is then interpolated to fill any remaining gaps.

5.3.3   Calculating Conversion Scales

By constructing the epipolar images as in the DPP method, the relative orientations and locations of the projected images should be preserved. This allows constructing proportionality between both the co-related disparities and their corresponding elevation above the selected projection plane. Additionally, the offset information calculated from the DPP method allows computing three transformation forms: (1) disparity-to-disparity scaling formula as derived in the DPP method, (2) disparity-to-elevation conversion formula (DEF), and (3) correct distance formula (CDF) to calculate the correct orthographic location in the epipolar direction.

Before providing the DEF and CDF required in the DECR method, the disparity value needs to be explained. The disparity maps represent the co-relation of one epipolar stereo image (reference image) with respect to another stereo mate (input image). Hence, the disparity value (D_{LR}), as expressed in **Equation (5.1)**, is simply the X' coordinate value of a pixel in the left epipolar image ($X^{L'}$) minus the X' coordinate value of the same pixel in the right epipolar image ($X^{R'}$). The resulting disparity map fits exactly the reference epipolar image with pixel-level co-registration.

$$D_{LR} = X^{L'} - X^{R'} \tag{5.1}$$

To derive the required conversion relationship, a descriptive sketch needs to be developed. **Figure 5.2** illustrates the relationship of the light rays projected onto the two horizontal object-space planes for the case of three MVIS-VHR satellite images. These

150

light rays are all represented in the epipolar plane–XZ (the page plane)–along the same epipolar line (i.e., the same row of the projected images). The intersection point of these light rays is the actual point imaged in the object space. The photogrammetric triangulation approaches solve for this point (space intersection) to reconstruct the original imaged object space. The intersections of the rays with the horizontal average level plane (Z_{avg}) are the ground locations of the projected pixels. Hence, the object-space distance between these projected pixels are the object-space disparities (D).

From the sketch shown in **Figure 5.2**, the disparity-to-elevation formula (DEF) can be derived based on the basic triangle similarity rules. This formula is based on scaling the object-space disparities as expressed in **Equation (5.2)**.

$$Elv. = \left(\frac{\Delta Z}{\Delta X^{m\prime} - \Delta X^{n\prime}}\right).D_{m.n} = \left(\frac{\Delta Z}{\Delta X^{k\prime} - \Delta X^{n\prime}}\right).D_{k.n} \qquad (5.2)$$

where $Elv.$ is the disparity-based derived elevation value above the selected projection plane; ΔZ is the difference between the selected two parallel object-space planes (in our case $Z_0 = 0$, then $\Delta Z = Z_{avg}$). Thus, the absolute elevation value from the zero level equals the calculated $Elv.$ value in addition to the elevation of the first projection plane (i.e., $Elv. + Z_{avg}$).
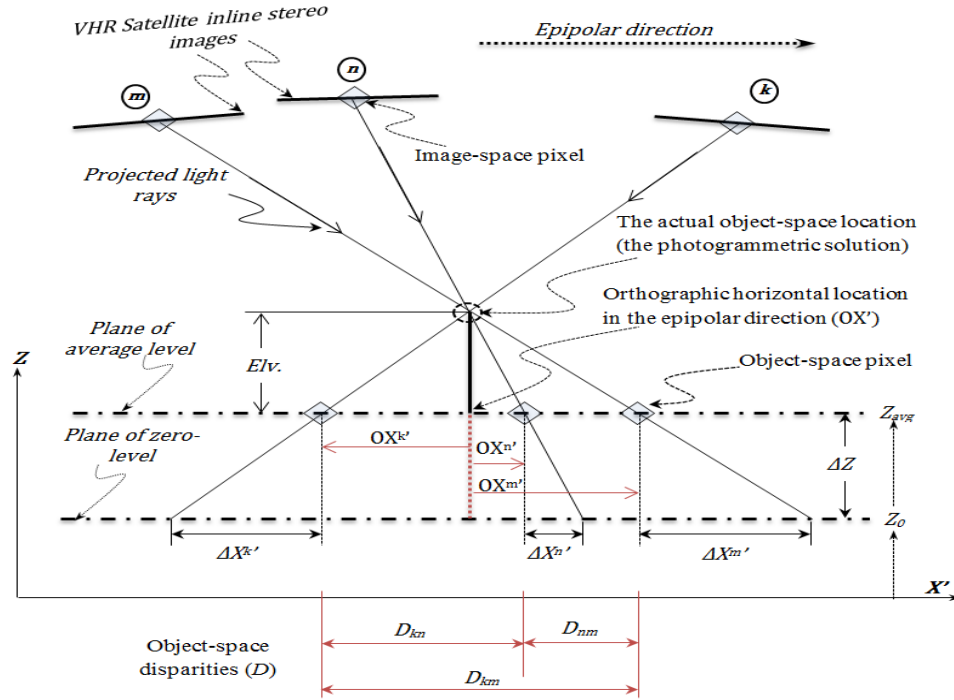
Figure 5.2 A sketch in the epipolar plane describing the relationships among the ground disparities and with their corresponding elevation above the selected projection plane.

From the same sketch in **Figure 5.2**, the correct distance formula (CDF) for calculating the shift of the ground pixels from the correct orthographic location in the epipolar direction ($\Delta X'_{ortho}$) can be found by the same manner. This value, as expressed in **Equation (5.3)**, can be calculated for each ground pixel based on its elevation value to give the correct orthographic location in the constructed epipolar geometry (X'_{ortho}).

$$\Delta X^{k'}_{ortho} = \left(\frac{Elv.}{\Delta Z}\right).\Delta X^{k'} \quad \rightarrow \quad X^{k'}_{ortho} = X^{k'} - \Delta X^{k'}_{ortho} \tag{5.3}$$

where $\Delta X^{k'}_{ortho}$ is the distance from the ground pixel ($X^{k'}$) of the epipolar image (k) to its correct orthographic location ($X^{k'}_{ortho}$) in the epipolar direction. Similar to that as indicated in **Figure 5.2** the distances for the projected image (m) and (n) are $\Delta X^{m'}_{ortho}$ and $\Delta X^{n'}_{ortho}$ respectively.

152

5.3.4 Deriving Co-registered Elevations

Once the enriched disparity map is generated, the resulting SDM, as in the second step, is already co-registered accurately to the selected reference image. Since the DEF scale has been calculated as in **Equation (5.2)** for the reference image, the enhanced SDM is then converted into elevation data that already fit exactly the reference image as per the disparity definition described earlier. Therefore, this will result in a LoS-DSM that is generated based on disparity data and co-registered with pixel-level accuracy to the relevant reference image selected from the MVIS-VHR satellite images.

The achieved LoS-DSM based on our DECR algorithm should be equivalent to the original algorithm with some additional advantages. For instance, since the employed images are projected into the object-space at the average terrain level of the imaged area, these images have minimal terrain relief effects. The distances in the epipolar direction between the corresponding terrain-level features in the constructed epipolar images should be the minimum. Therefore, the measured disparities of the terrain-level features in the generated SDMs should be close to zero (from the projection plane level) in the areas of moderate terrain-relief variation as demonstrated in Suliman et al. (2016).

Since the enriched SDM is generated with minimized terrain-relief effects, the derived corresponding elevations are of minimal terrain-relief variation and hence approximately normalized to represent the off-terrain features. Consequently, the product of this step is the LoS-DSM. This LoS-DSM is efficiently derived from scale-related disparity data, co-registered with pixel-level accuracy to the reference epipolar image, and has minimized terrain relief effect

5.4 DECR Method Validation

The proposed validation procedure for the DECR method includes three validation steps. These steps are (1) confirming the epipolarity and validating the disparity proportionality among the employed VHR images, (2) validating quantitatively the disparity-based derived elevations, and (3) inspecting qualitatively the co-registered LoS-DSM. Further details are provided in the following subsections.

5.4.1 Epipolarity and Inter-proportionality Confirmation

The confirmation of the epipolarity of multiple epipolar images can be carried out qualitatively by arranging these images horizontally and checking whether the same image point in all of the constructed epipolar images can be connected by a straight horizontal line or not.

Quantitatively, the epipolarity condition is validated by applying an automatic matching technique to find a set of accurate point matches and calculate the differences in the y-disparity (y-parallax) for all point pairs. For this check, the scale-invariant-feature-transform (SIFT) matching technique introduced by Lowe (2004) is selected due to its accuracy and reliability.

The validation of the computed epipolar angle and the offsets in the epipolar direction ($\Delta X'$) is indicated by the range of the offsets ($\Delta Y'$) perpendicular to the epipolar direction. This range should be within one pixel.

However, as introduced in Suliman and Zhang (2016), to validate the disparity inter-proportionality among the constructed epipolar images, the same image point and its corresponding ones should lie on the same straight line if the images are arranged

vertically by distances proportional to the calculated offsets in the epipolar direction. **Figure 5.3** illustrates this arrangement (but in horizontal manner) that satisfies **Equation (5.4)** which represents the validation line slope.

$$\frac{L_{no}}{L_{nm}} = \frac{D_{no}}{D_{nm}} = \left(\frac{\Delta X^{n\prime} - \Delta X^{o\prime}}{\Delta X^{n\prime} - \Delta X^{m\prime}}\right) \tag{5.4}$$

where the L_{no} is the distance between the corresponding epipolar lines in the epipolar image (n) and image (o); D_{no} is the object-space disparity for the same point between the epipolar image (n) and image (o); likewise for L_{nm} and D_{nm} respectively. This ratio can be determined from the calculated offset information in the epipolar direction (i.e., $\Delta X\prime$) for each image, as described in the DPP method. By knowing this ratio, the intermediate distances (Δ_{ij}) between the consecutive images ($\Delta_{image1, image2}$) can be determined based on basic triangle similarity rules after selecting the reference image and the reference stereo-pair domain. All of these symbols are illustrated in **Figure 5.3**.
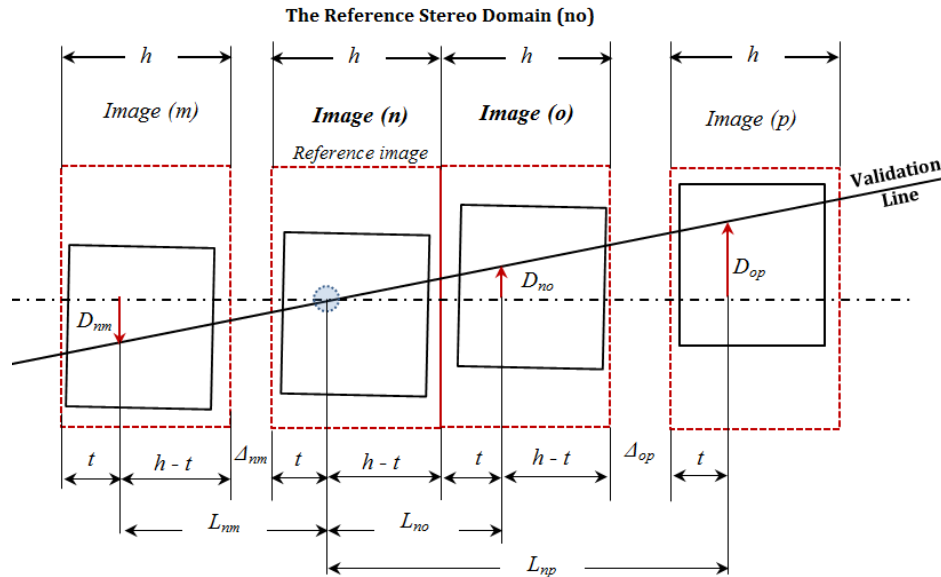
Figure 5.3 The inter-proportionality validation for the object-space disparity values from different stereo domains

155

5.4.2   Calculated Elevations' Validation

After constructing the epipolar images of scale-related disparities to their corresponding elevation and deriving the DEF scale value, elevation values can be derived from the measured disparities for all matched pixels in the reference image. These elevation values should be equivalent to those derived photogrammetrically. Hence, as an external reference to evaluate the proposed disparity-based elevation calculation, the DSM for the imaged area needs to be generated by the traditional photogrammetric approach which is already geo-referenced. However, a direct comparison for the derived elevations cannot be straightforwardly made. This is due to the projection difference between the two sets of elevation data. While the photogrammetric DSM has an orthographic and geo-referenced projection, the developed disparity-based LoS-DSM has a perspective projection (line-of-sight) that is not geo-referenced.

Therefore, for this elevation validation, a set of well-distributed points in the generated LoS-DSM need to be selected and corrected for their orthographic locations to be compared against those derived photogrammetrically. As a simplified technique, a set of points that represents the centers of flat rooftops of buildings in the generated LoS-DSM can be selected and matched manually to their corresponding ones in the photogrammetrically-derived DSM. Hence the corresponding elevations can be compared against each other. This technique is recommended to bypass the need to co-register the points in the perspective LoS-DSM with their corresponding ones in the orthographic DSM.

In contrast, as a more accurate technique, the set of well-distributed points selected in the generated LoS-DSM need to be corrected for their orthographic locations by calculating the $(\Delta X'_{ortho})$ for each point by the CDF formula expressed in **Equation (5.3)**. Then, the new point distributed is rotated in reverse of the rotation angle applied to construct the epipolar images. Consequently, the corresponding elevations in both disparity-based LoS-DSM and the photogrammetric-based DSM are co-registered and then can be compared.

Once the corresponding elevations from the two elevation models are co-registered, the elevation differences can be calculated, and hence statistical quality measures can be computed such as the root-mean-square-error (RMSE) and the standard deviation (Std.) of the resulting discrepancies.

5.4.3   Co-registered LoS-DSM Validation

After generating the enriched SDM and deriving the LoS-DSM co-registered to the optical reference image, the homogeneity of the values related to the same object can be identified qualitatively through a 3D rendered representation of the generated LoS-DSM and validated visually. This inspection allows creating isometric views of the optical data that are helpful in evaluating the quality of the elevation data generation and co-registration.

Once the geometric and optical data are integrated, and a complete set of data ideal for 3D-based information extraction applications are achieved, this set of data is required to be incorporated in a mapping application for a dense urban area to demonstrate its applicability. The following section describes the proposed procedure for this application.

5.5   Applicability Demonstration

Co-registering optical and elevation data leads to an ideal dataset for 3D-assisted information extraction applications. In this section, elevation-based building detection (EBD) in VHR satellite imagery is selected to be implemented for demonstrating the applicability of the disparity-based elevations co-registered through the improved solution of generating LoS-DSM (i.e., DECR method). In contrast to the process demonstrated in Suliman and Zhang (2015), executing the EBD application in our study is meant to validate the terrain-relief minimization by representing approximately the off-terrain objects only. Furthermore, it includes geo-referencing the finished building-roof objects. The proposed EBD and geo-referencing process are described in the following subsection.

5.5.1   Elevation-based building Detection (EBD) and Geo-referencing

The elevation-based building detection method described in Suliman and Zhang (2015) is selected and adapted for our study to demonstrate the applicability of the co-registered disparity-based LoS-DSM. The steps of the procedure start by pan-sharpening the reference image for better classification and segmentation results. Then, the developed DECR method is executed. Once the optical-elevation data co-registration is achieved, the optical data are segmented and the vegetation objects are suppressed. After that, building objects can be detected easily using an elevation threshold value. After enhancing the result, finally, the detected building objects are geo-referenced to their correct geo-locations. All of these steps are flowcharted in **Figure 5.4**.
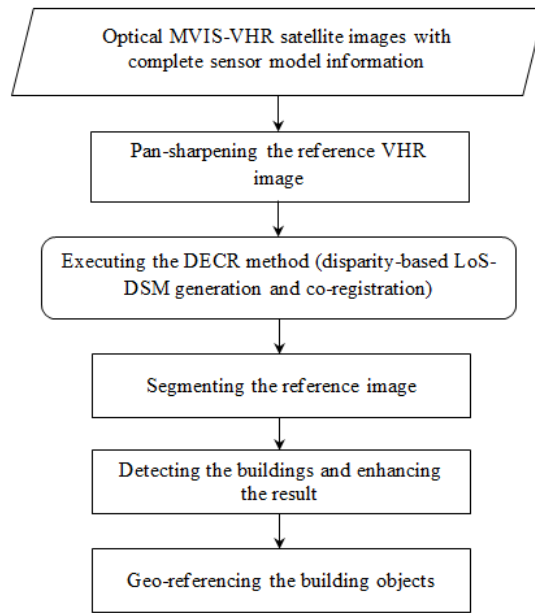
Figure 5.4 The proposed elevation-based building detection (EBD) and geo-referencing procedure.

5.5.1.1 Input image pan-sharpening

VHR images have one panchromatic band and at least four multispectral bands (Red, Green, Blue, and Near-Infrared). The ground resolutions of the multispectral bands are usually one-fourth of that of the panchromatic band. Therefore, to combine these bands and take the advantages of both types, an image fusion technique needs to be executed. The UNB pan-sharpening technique introduced by Zhang (2004) is highly recommended. We propose executing this fusion technique just for the image selected for applying building detection.

5.5.1.2 Disparity-based LoS-DSM generation

For achieving co-registered elevation data to the reference VHR optical image, the disparity-based LoS-DSM needs to be generated. Therefore, all of the steps described in

Section 5.3 for the DECR method are required to be executed to create a complete integrated dataset through the optical-elevation data co-registration.

5.5.1.3 Image segmentation

To reduce the image complexity and divide it into small objects based on a homogeneity measure of the color information, an image segmentation technique should be executed. As concluded by Dey (2013), multi-resolution segmentation introduced by Baatz and Schäpe (2000) is one of the most appropriate techniques for segmenting VHR images of urban areas. Thus, this technique is adopted in our study for the segmentation step.

5.5.1.4 Building detection and results enhancement

To detect building objects, a thresholding operation for the achieved normalized elevations needs to be applied. However, since the elevations are going to have a crucial detection role, other elevated features, such as trees, must be removed to avoid confusion with the building objects. Fortunately, vegetation indices can be used effectively to detect and suppress vegetation. The normalized deference vegetation index (NDVI), as described in Teillet (1992), is one of the most accurate and popular vegetation indices. In our study, this index is selected to suppress the scene vegetation. After suppressing the vegetation, the building objects can be detected directly by a applying an elevation threshold. The best thresholding value can be selected empirically.

Since the off-nadir VHR images have severe building lean, an occlusion detection technique should be executed to detect the hidden areas which usually represent the building sides (façades) and create gaps and outliers. The most common techniques of

160

detecting the occluded areas are reviewed by Egnal and Wildes (2002). We recommend the use of the Left-Right checking technique due to its efficiency and simplicity. Masking out the detected occlusion areas should highlight the building rooftops and hence enhance the result significantly.

Normally, the detection result requires some finishing. Therefore, post-processing procedures need to be applied for the resulting roof objects. These procedures are usually based on morphological operations. We recommend developing a rule set of such operations interactively and effectively using the Cognition Network Language (CNL) available with the *eCognition®* software package. To assess the result accuracy, the conventional evaluation measures described in **Section 5.5.2.1** are used to evaluate the building detection accuracy.

5.5.1.5 Building objects' geo-referencing

After detecting the building objects and enhancing their shapes, these objects should be geo-referenced to their correct orthographic locations. This allows easy input and integration with other existing GIS layers. Therefore, these objects need to be moved and registered to their correct geo-locations.

To determine the correct geo-referencing location, two steps are required: (1) moving these objects to the orthographic location in the epipolar direction as expressed in the CDF formula, and (2) rotating these objects to the reverse of the epipolar rotation direction (i.e., $-\theta$) applied in the epipolar construction phase. To perform the first step, the building objects are abstracted by representative points (RPs) that lie within their corresponding object's boundary. To guarantee that this RP is located inside the

corresponding object's shape, this point is selected to represent the center of the greatest circle that fits inside the object boundary. The algorithm for calculating this RP point location is introduced by Garcia-Castellanos and Lombardo (2007). By executing these two easy steps, the resulting building map should be geo-referenced to the correct orthographic geo-location. The validation test of the geo-referencing result is described in **Section 5.5.2.2**.

5.5.2   Mapping Accuracy Evaluation

This section describes the accuracy assessment and evaluation for both the detection and geo-referencing results. Once the building-roof objects are detected, the accuracy is assessed based on reference data using the traditional measures of the detection performance. Then, after geo-referencing the mapped roofs, the achieved result is evaluated.

5.5.2.1 EBD accuracy assessment

The detection accuracy is assessed by comparing the mapped buildings' roofs against a reference dataset after implementing the developed EBD procedure. Completeness, Correctness, and Overall Quality are three widely used measures to assess the detection performance as applied in Suliman and Zhang (2015). These assessment measures are a modified version of those introduced by Story and Congalton (1986). Completeness is the percentage of the entities in the reference data that are correctly detected. Correctness indicates how well the detected entities match the reference data. The Overall Quality of the results provides a compound performance metric that balances completeness and

correctness (Rutzinger, Rottensteiner, & Pfeifer, 2009). The formulas of these measures are as follows:

$$Completeness\ (Comp.)\ = \frac{\text{TP}}{TP+FN} \tag{5.5}$$

$$Correctness\ (Corr.)\ = \frac{\text{TP}}{TP+FP} \tag{5.6}$$

$$Overall\ Quality\ (OQ)\ = \frac{\text{TP}}{TP+FP+FN} \tag{5.7}$$

True positive (TP) is the number of correctly identified building roof segments. The false negative (FN) is the number of building roof segments in the reference dataset that are not detected or wrongly labeled as not roofs. The false positive (FP) represents the number of building roof segments that are detected but do not correspond to the reference dataset. It is worth mentioning that the detection assessment can be pixel-based or object-based (Rottensteiner, Trinder, Clode, & Kubik, 2005; Rutzinger et al., 2009). However, these measured entities in this study to represent the total number of pixels that are related to the building-roof class.

5.5.2.2   EBD geo-referencing evaluation

In many cases, the accurate reference maps for the building footprints are not available for the study area. Therefore, for validation purposes, we propose the generation of an orthophoto by using the traditional differential orthorectification technique. This technique requires the elevation data to be available in advance. Hence, the geo-referenced DSM should be derived photogrammetrically for the study area before executing the orthorectification process. The generated orthophoto will be used as an external reference data source to validate and assess the quality of the geo-referenced

building roof objects. This validation is proposed to visually inspect geo-referenced building roofs individually (object by object) in order to count the correctly geo-referenced objects ($\geq$ 90% overlap with the correct rooftop). This is because the traditional orthorectification process does not work perfectly in dense urban areas. Although true orthorectification overcomes the limitation of the traditional orthorectification process, the generation of a true orthophoto is expensive, time consuming, and difficult to achieve (Suliman & Zhang, 2015).

5.6 Data, Results, and Discussion

The results of the developed DECR method and its demonstrated applicability in EBD application are presented and discussed in this section. The section organization is based on the sequence of the proposed validation procedures described in **Section 5.4** and **Section 5.5.2** for the DECR and the EBD methods respectively. The test data used in the experimentation are described in the following subsection.

5.6.1 Dataset and Study Area

The optical data used in this research are a subset of five stereo VHR images acquired by the WorldView-2 linear sensor with push-broom scanning mode. These MVIS images are available along with their sensor model information. Each of the five MIS-VHR satellite images has eight spectral bands of 1000$\times$1000 pixels and a panchromatic band of 4000$\times$4000 pixels with 2m and 0.5m resolutions, respectively.

The VHR data were acquired in 2010 over a dense urban area of Rio de Janeiro, Brazil. The imaged area is of a modern city and has many buildings with different shapes and sizes. Most of the buildings are high-rise elevations where building lean and façades

are prominent. **Figure 5.5** shows the selected reference stereo pair (i.e., the reference image and its stereo mate).

<div align="center">(a)   (b)</div>
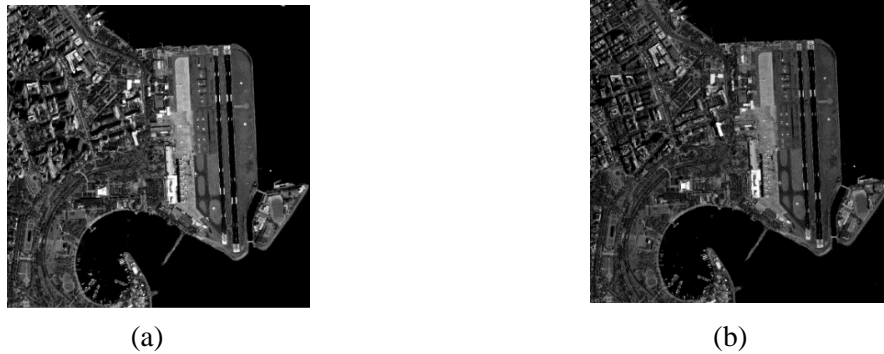
Figure 5.5 The reference stereo pair of the MVIS-VHR satellite images. (a) The reference image (I-2). (b) The stereo mate (I-3).

5.6.2 Results of DECR Method

5.6.2.1 Epipolarity and inter-proportionality result

Based on the developed DECR method, the available VHR images of the test data are re-projected onto two horizontal planes at the average-terrain elevation (i.e., $Z_{avg} = 10m$) and the zero elevation (i.e., $Z_0 = 0m$). To simplify the calculations, the resampling of the re-projected images was made to be equivalent to $1m^2$ per pixel. After that, the shift information in the X and Y direction of the object-space coordinate system was calculated for each MVIS-VHR image, as shown in **Table 5.1**.

To confirm the epipolarity condition of the epipolar rectified images, the SIFT matching technique was executed to generate a set of 35 matching points of good distribution. Then the differences in the matching location were calculated between the point pairs to assess the quality of the epipolarity condition. The RMSE value of these differences was found to be 0.29 pixels with a Std. value of 0.23 pixels. These values

validate the approximation of the epipolar curves and indicate the success of the epipolar rectification process for the employed MVIS-VHR satellite images.

Table 5.1 shows the minimum, maximum, and the average shift values in the X and Y directions respectively. It can be noted that the variation of the values is very small ($<$ 1dm) due, mostly, to the small size of the test area. In this case the average values can be used as a generalized value for the whole test area–to simplify the calculations–instead of using the individual values of each ground pixel.

Table 5.1 The calculated offset information required for the DECR method.

Original Image	I-1	I-2	I-3	I-4	I-5
$\Delta X_{min.}$	-2.3113	-1.5172	-0.2792	1.4300	2.4262
$\Delta X_{max.}$	-2.2759	-1.4822	-0.2436	1.4646	2.4610
ΔX_{avg}	-2.2937	-1.4997	-0.2615	1.4473	2.4435
$\Delta Y_{min.}$	-9.8337	-6.5930	-1.5081	5.6205	9.8155
$\Delta Y_{max.}$	-9.7840	-6.5436	-1.4566	5.6363	9.8406
ΔY_{avg}	-9.8091	-6.5685	-1.4824	5.6283	9.8277

[*]The values are in meter (ground pixel size = 1m)

To align the rows of the images with the epipolar direction, the rotation angle of the epipolar direction was determined as described in the DPP method. Since the imaged area was of small size, the rotation angle can be computed from the average values in **Table 5.1** of any stereo pair. To ensure that this angle is approximately constant regardless of the stereo pair used to calculate it, the epipolar direction is computed based on the average offset values for four different combinations of stereo pairs. The calculated angles are almost equal regardless of the stereo pair used. As a result, the most representative rotation angle is the average value. Thus, for this dataset, the rotation to the epipolar direction was found to be 76.4231 degrees.

To construct the epipolar images and keep their inter-relationships, the projected images were rotated using the calculated epipolar angle. Additionally, the calculated offset data were rotated based on the same angle to find the corresponding offsets in the epipolar direction and its perpendicular (i.e., $\Delta X'$, $\Delta Y'$). The offset values in this direction were all found to be less than 1 pixel (or 1 meter) with a maximum range of about 0.75 pixels. This indicates quantitatively the validation of the computed epipolar angle, the calculated offsets in the epipolar direction, and the approximated epipolar lines of the constructed epipolar images. This validation is demonstrated visually in **Figure 5.6**. The same point and its corresponding points lie on the same horizontal line in all epipolar images.
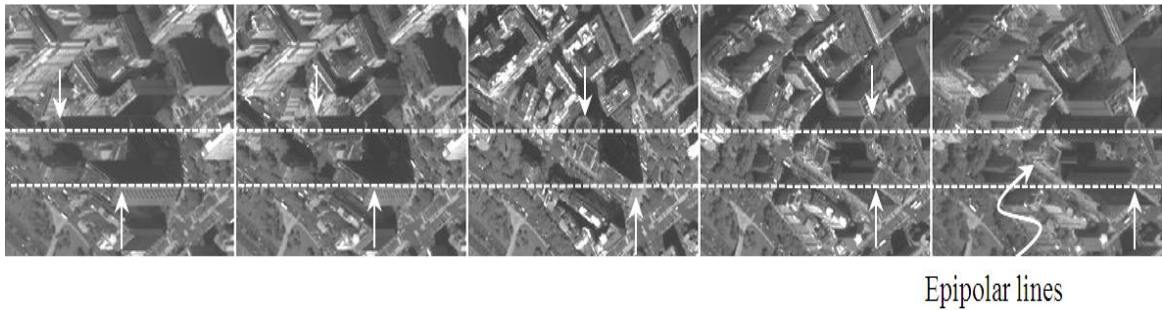
Epipolar lines

Figure 5.6 Epipolarity validation. Any point and its corresponding ones lie on the same horizontal line in all epipolar images

To validate the inter-proportionality of the disparity values among all of the epipolar images, the validation test described in **Figure 5.3** was implemented. **Figure 5.7** is constructed based on **Equation (5.4)**. The figure proves the validity of the disparity proportionality among all epipolar images. Two straight lines of two different points indicate the success of the constructed disparity proportionality. This validation confirms the proportionality of the disparity values with their corresponding elevations.
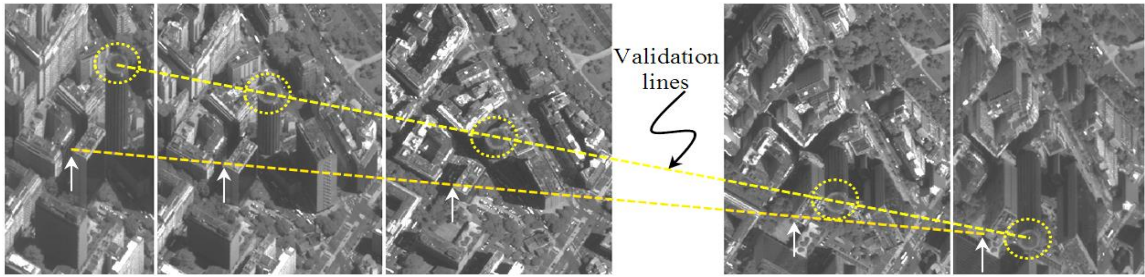
Figure 5.7 Disparity inter-proportionality validation among all epipolar images

5.6.2.2 Calculated elevations result

Since the scale-relationship of the disparity values was achieved and validated, supplementary surface disparity data (SDMs) were generated as described in the DPP method. These supplementary disparity data were used to fill disparity gaps of the reference disparity map. By fusing the generated SDMs with the reference one, the SDM of the selected reference pair (i.e., Ep-2&Ep-3) was enriched. **Figure 5.8** shows different SDMs extracted from different epipolar stereo pairs which are fused together to generate a more accurate and enriched SDM. The enriched SDM achieved has an identical representation of its corresponding LoS-DSM.

SDM of (Ep-2 &Ep-1) stereo domain   SDM of (Ep-2 &Ep-3) stereo domain   SDM of (Ep-2 &Ep-4) stereo domain   Enriched SDM of (Ep-2 &Ep-3) stereo domain
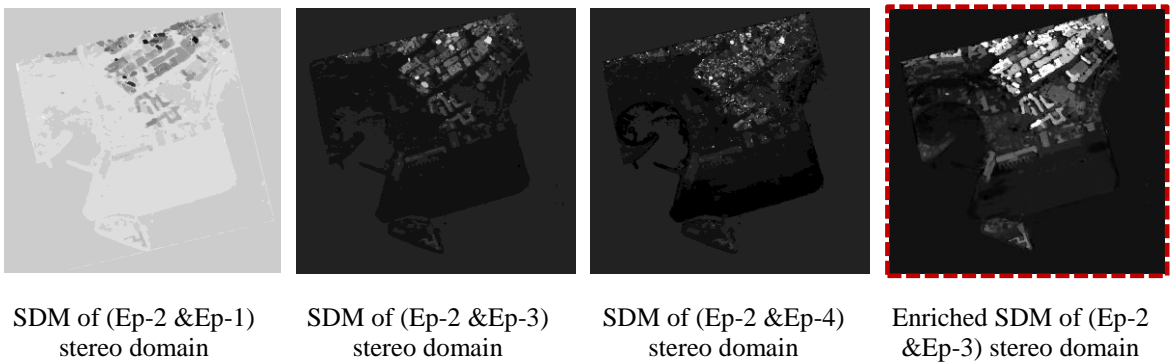
Figure 5.8 The enriched SDM generated from different epipolar stereo pairs based on the scale transformation formula derived in the DPP method. Unlike the rest of the pairs, the pair Ep-2& Ep-1 has negative values because the stereo mate is before the reference image in the sequence of acquisition.

168

As stated previously, the shift variation is almost negligible due to the small size of the test area. Thus, to simplify the calculations, the average shift of each image in the epipolar direction was used to calculate the DEF formula –as in **Equation (5.2)** – instead of the shift value for each individual pixel which must be considered in the case of full scenes and hilly areas.

Based on **Equation (5.2)**, the scale value calculated for the selected reference stereo pair (I2-I3) is found to be -1.912. Hence, the elevation data corresponding to the enriched SDM of the reference stereo domain were directly generated for the selected reference stereo pair (I2-I3). Since these elevations are measured from the projection plane (Z_{avg}), the height above the datum of that projection plane ($\Delta Z = 10\text{m}$) is added to have the elevation referenced to the datum. Hence, the disparity-based derived elevations are computed as $Elv. = 10 - 1.912 \times D_{I2,I3}$. The derived elevation data represent the disparity-based LoS-DSM co-registered to the reference image with pixel-level accuracy.

For the evaluation of the disparity-based elevation calculation, a set of tie points that represent the centers of different flat building-rooftops were matched manually in order to correctly calculate their 3D ground coordinates. The resulting ground elevations were then compared against their corresponding values generated based on the DECR method after geo-referencing them as described in **Section 5.5.1.5**. The photogrammetric elevations were calculated using industry leading commercial photogrammetric software (PCI Geomatica, version 2015). Based on 10 building roof elevations, which were selected manually, the RMSE and Std. were calculated.

169

The RMSE value was found to be less than 2 pixels (1.54m). This value is reasonable because it is less than the threshold value (i.e., 2 pixels) used to identify the inconsistency among the corresponding supplementary disparity values generated from different epipolar stereo pairs before applying the merging technique (used in the DPP method) as described in Suliman and Zhang (2016). Moreover, this threshold is almost double the standard deviation value (1.06m). Hence, this indicates the highly acceptable precision of the achieved RMSE value. Therefore, the developed disparity-based approach is able to produce elevation data comparable in accuracy to that derived rigorously based on the traditional photogrammetric approaches using industry leading commercial software.

5.6.2.3 Co-registered LoS-DSM result

Based on the DECR method, the resulting LoS-DSM co-registered to the selected reference image needs to be validated to identify the elevation inconsistency through visual inspection.

(a)   (b)
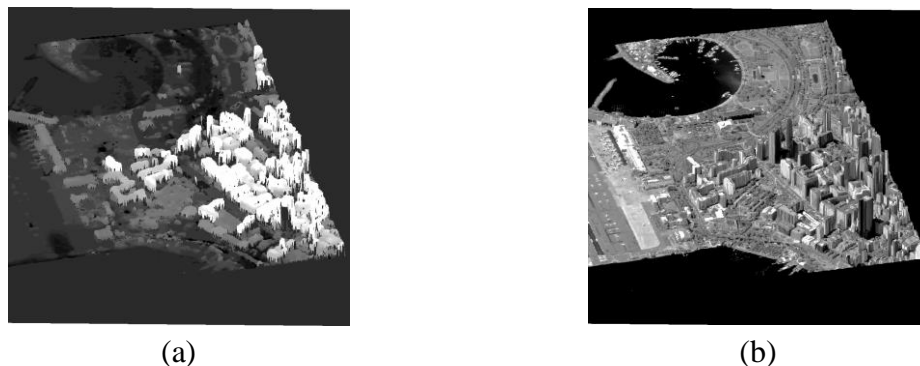
Figure 5.9 The co-registered LoS-DSM validation (a) 3D isometric view of the generated LoS-DSM based on the DECR method. (b) 3D rendered representation of the LoS-DSM co-registered to the selected reference image.

Figure 5.9 shows an isometric view of the generated LoS-DSM and the rendered representation of the co-registered reference image. While **Figure 5.9-(a)** shows clearly the elevation consistency for each building roof, **Figure 5.9-(b)** indicates the success of the co-registration with the reference optical data. The realistic 3D representation of the dense urban area is attributed to the accurate optical-elevation data co-registration achieved using the developed DECR method.

5.6.3   Results of EBD Procedure

To demonstrate the applicability of the developed DECR method, an elevation-based building detection application was implemented to map the building-roof objects based on the accuracy of the elevation derivation, terrain elevation minimization, and the pixel-level co-registration with the reference optical image.

The building detection is mainly based on the elevation information. Because trees produce false detection results, a mask was generated based on the NDVI value to exclude all of the vegetation objects. The threshold value of this index was selected empirically and was equal to 0.3. It is worth mentioning that this NDVI-based vegetation removal may affect negatively the detection result in the cases of roof gardens or buildings with high NDVI values. These two cases pose a limitation to this vegetation removal technique. However, these cases rarely appear in most urban areas.

Since the reference image was selected to be off-nadir over a challenging and dense urban area, the buildings' façades are prominent. To exclude these confusing building objects, the Left-Right occlusion detection technique was executed to identify the hidden areas which mainly represent in our case the building façades.

Figure 5.10 illustrates an example of the performance of this technique (**Figure 5.10-(c)**). The whole façade bitmap generated for the selected epipolar stereo images (Ep-2 & Ep-3) is provided in **Figure 5.10-(d)**. This mask is used to enhance the building detection results by removing the building sides and highlighting the building rooftop objects only.

(a)　　　　　　　(b)　　　　　　　(c)　　　　　　　(d)
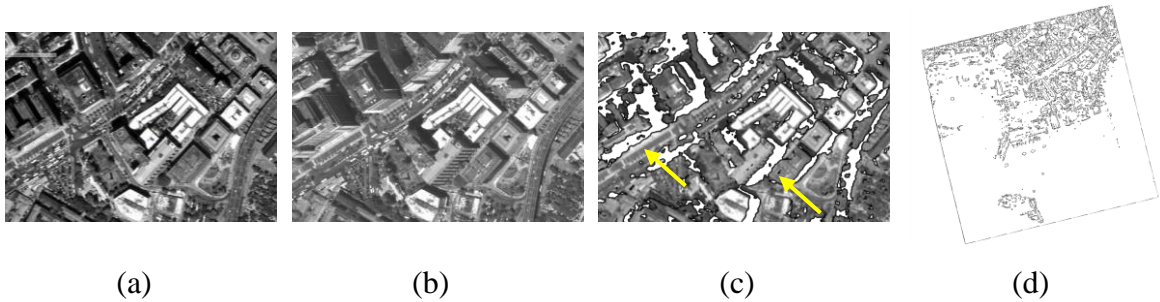
Figure 5.10 The generation of the occlusion map for the reference stereo images (I-2&I-3). (a) A subset from the epipolar referenace image (Ep-2). (b) A subset from the epipolar stereo mate image (Ep-3). (c) The detected occlusions in the reference image. (d) the generated occlusion map for the epipolar reference image (Ep2)

The intermediate detection results are provided in **Figure 5.11**. The co-registered optical and elevation datasets are shown in **Figure 5.11-(a)** and **Figure 5.11-(b)** respectively. **Figure 5.11-(c)** demonstrates the detected off-terrain objects based on a thresholding operation. The threshold value was very close to the selected terrain average elevation (Z_{avg}) because the elevations were referenced to the datum. This indicates the success of the terrain variation minimization described in the developed DECR method. The detection result of **Figure 5.11-(c)** includes the tree objects which are filtered out in **Figure 5.11-(d)** based on an NDVI bitmap produced based on the empirically selected threshold value.

172

This result is enhanced further by applying the Left-Right-based occlusion map of **Figure 5.10-(d)** generated to exclude the building façades. Some post-processing steps were applied to enhance the detection result. These steps include merging adjacent objects and removing isolated ones of small areas since they usually represent noise. The final enhanced detection result is shown in **Figure 5.11-(e)**. The reference data used for evaluating the detection results is provided in **Figure 5.11-(f)**. These data were generated manually. The quantitative evaluation of the final detection result relative to the reference data is provided in **Table 5.2**.

Table 5.2 Performance evaluation of the building detection result.

Evaluation measure	Comp. (%)	Corr. (%)	OQ. (%)
Value	96	95	**92**

Based on the performance measure provided in **Table 5.2**, the detection was highly successful. This high quality detection is due to the incorporation of the elevation information. This information is a critical detection component for the buildings since they are inherently elevated objects.

The high correctness value of 95% is attributed mainly to the high accuracy of the derived elevation data used for the detection and the accurate co-registration of these elevation data with the reference image. This performance indicates the success of the developed DECR method. Furthermore, the quality of the elevation data generation and co-registration is reflected also in the high Completeness of 96%. However, the role of incorporating the façades bitmap based on the Left-Right checking technique reduced the false detection tremendously which resulted in a higher Completeness measure value.

As a combined indicator of both the correctness and Completeness performance measures, the Overall-Quality measure confirms the high quality of the building detection
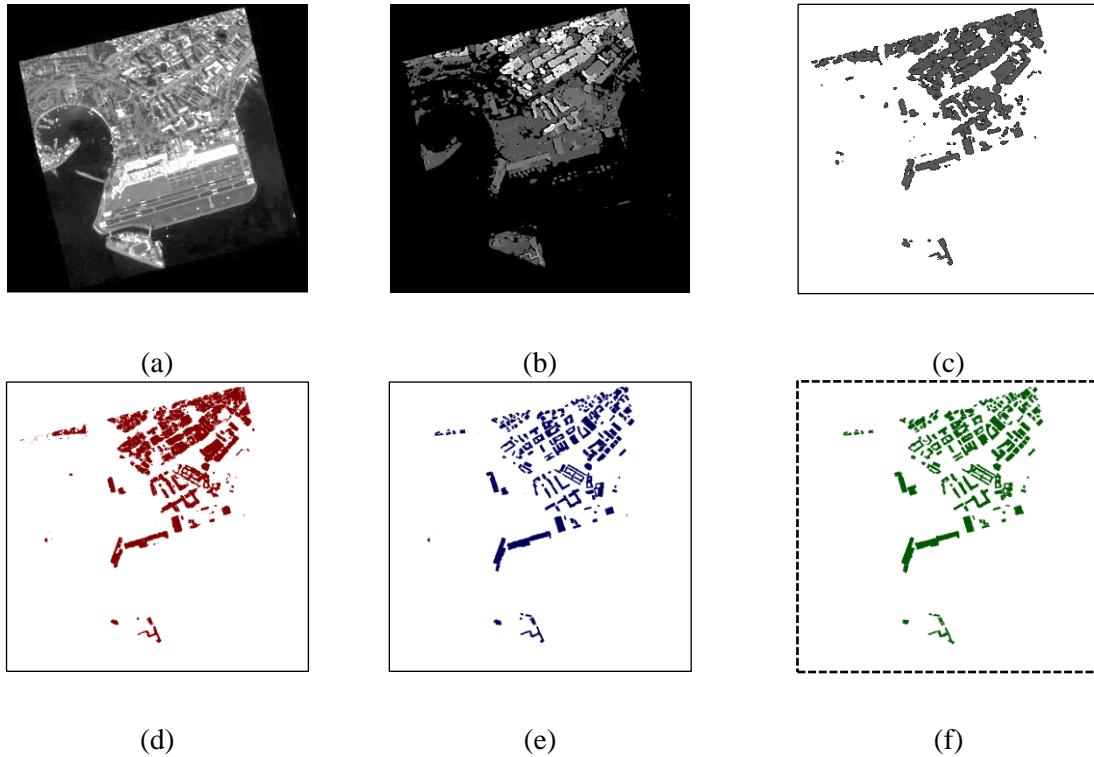
Figure 5.11 The intermediate building detection results. (a) Epipolar off-nadir VHR reference image (Ep-2). (b) Disparity-based co-registered LoS-DSM using DECR method. (c) Detected off-terrain objects based on a thresholding operation of a value close to Z_{avg} . (d) Resulting objects after suppressing vegetation objects based on an NDVI bitmap. (e) Resulting objects after removing the building façades. (f) The manually generated reference data for comparison.

results. This quality can be seen by comparing visually the reference data in **Figure 5.11-(f)** to the final detection results in **Figure 5.11-(e)**. A building detection result in off-nadir VHR imagery of 92% Overall Quality over a challenging dense urban area is indeed a significant success that demonstrates the applicability of the developed DECR method in an EBD application.

174

5.6.4 Results of EBD Map Geo-referencing

After mapping the building-roof objects and evaluating the detection performance, the resulting building objects must be geo-referenced to their correct ground locations in order to be ready for incorporation in GIS systems and integration with other GIS layers. The geo-referencing process described in **Section 5.5.1.5** was implemented to remove the perspective effects of the off-nadir VHR image and then apply the geo-referencing. The correct geo-location in the epipolar direction was calculated based on **Equation (5.3)** using the derived elevations, referenced from the projection plane (Z_{avg}), at the RP point of each building object. For the selected reference image, the value used for ΔZ is 10m and for $\Delta X'$ is −6.7352m.

The achieved geo-referencing result was evaluated as described in **Section 5.5.2.2** based on an orthoimage. Therefore, a dense geo-referenced DSM was derived photogrammetrically based on the traditional approach and then used to create an orthoimage for the study area. This orthoimage was used as an external reference to assess the accuracy of the building map geo-referencing process.

Figure 5.12-(a) illustrates the detection result in the epipolar image domain overlaid on the epipolar reference image (I-2). **Figure 5.12-(b)** shows the geo-referencing result overlaid on the generated orthoimage. **Figure 5.12-(c)** shows the enhanced building detection results geo-referenced to the correct object-space locations and prepared as a GIS layer.

| (a) | (b) | (c) |
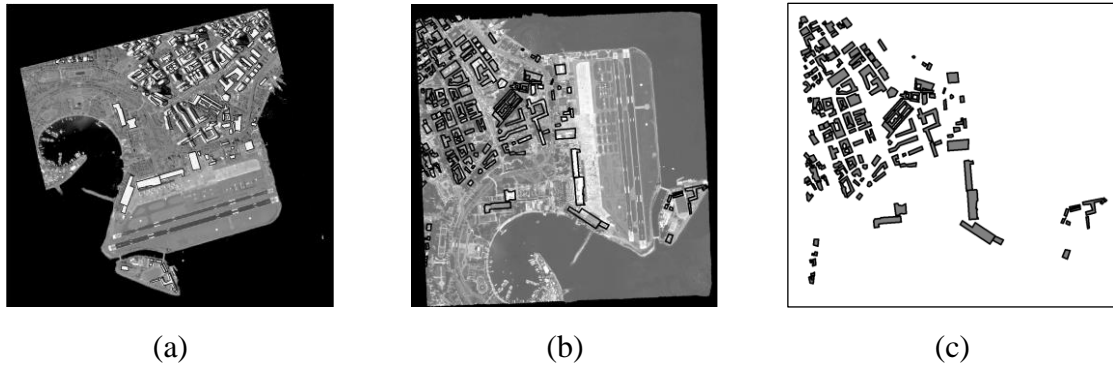
Figure 5.12 The result of the building detection in both the epipolar image domain and the geo-referenced ground domain. (a) The detected building objects in the epipolar reference image domain, (b) geo-referenced building objects overlaid on the generated orthoimage, and (c) the resulting GIS layer.

After a qualitative assessment of the generated orthoimage, it is found that 2% (3 out of 165 objects) of the geo-referenced building objects lie on the incorrect geo-location. The reason for that is the incorrect elevation used to calculate this geo-location in the epipolar direction. This elevation was extracted at the RP point of the building object as described early. For non-flat building roofs, the location of the RP selected in our study may not be at the elevation value that leads to the correct geo-reference location. These cases are very few and they can be easily identified and corrected manually.

Figure 5.13 illustrates four cases of the achieved geo-referencing results. On the one hand, the upper row of this figure shows the detection result in the epipolar image domain (Ep-2) along with the correct geo-location of the detected building object in the epipolar direction.

The shift of the detected objects to the correct location in the epipolar domain is equal to the distance between the RP point of the detected building (represented by a triangle in **Figure 5.13**) and correct building object location (represented by a square in **Figure 5.13**) as indicated in the same figure.

(a1)	(b1)	(c1)	(d1)

(a2)	(b2)	(c2)	(d2)

▲   The location of RP of detected roof object in the epipolar image.

▢   The location of the RP of the moved roof object to the correct location in the epipolar image.
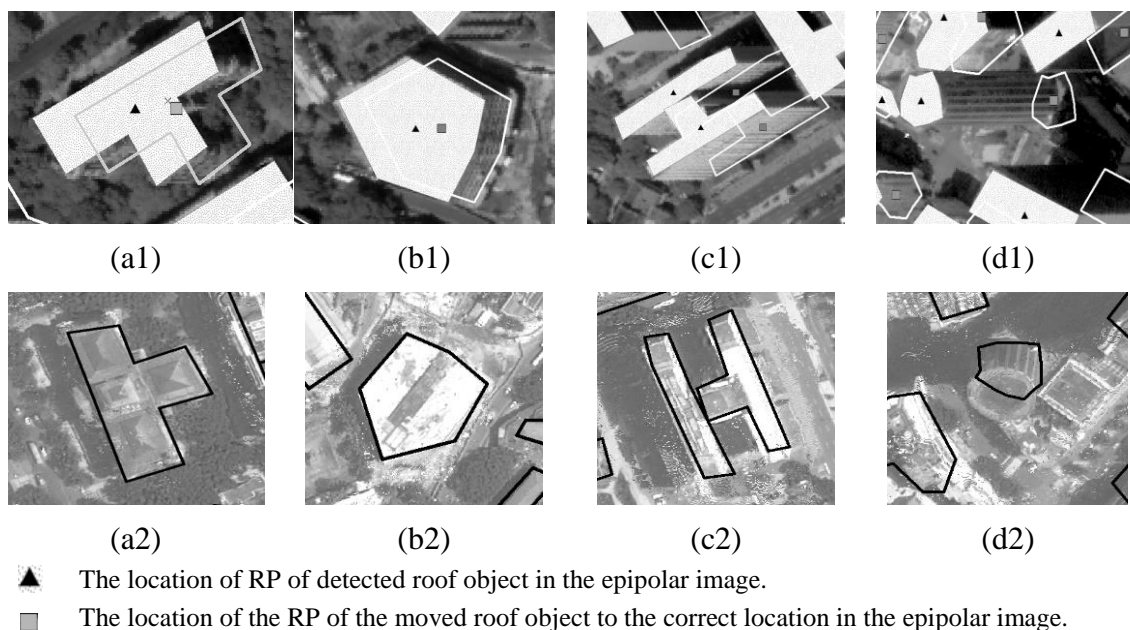
Figure 5.13 A few examples of the geo-referenced roof objects for visual assessment. The upper row shows the detected roof objects along with its RP location. The same row shows the calculated correct location of the object to be moved to it along with its new RP point.

On the other hand, the lower row of **Figure 5.13** shows the geo-referencing result overlaid on the generated orthoimage. Based on the visual assessment, the geo-referencing result is highly successful. For **Figure 5.13-(d2)**, despite the misalignment between the building roof in the orthoimage and the corresponding geo-referenced object, the location of the geo-referenced roof object is correct. This is because the lean of the high-rise building (> 135m above the local ground) was not completely eliminated from the orthoimage. Hence, the source of this misalignment is in the reference data, not the geo-referenced objects. **Figure 5.13** indicates the success of the proposed geo-referencing process. The final geo-referencing result was processed to be directly integrated into a

GIS system. At this stage, the cycle of the information extraction and mapping is complete for the application of building detection in off-nadir VHR images.

5.7 Conclusions

In this study, the problem of optical-elevation data co-registration is addressed. The introduced solution is based on generating disparity-based elevation data co-registered with one of the employed VHR images. The co-registration is achieved with pixel-level accuracy using an improved approach for developing a LoS-DSM from an enriched disparity map. This map is generated from different MVIS-VHR satellite images using the DPP method. By extending this method, the LoS-DSM elevations are derived efficiently based on the DECR method developed in this study.

The developed DECR method is based on establishing a scale-based relationship between the disparity and elevation data in the object space. This relationship is derived using the DPP method that calculates disparity scales for generating enriched disparity maps. The core concept of both DECR and DPP methods is to re-project the MVIS-VHR images onto an object-space horizontal plane at the average terrain elevation of the imaged area. This re-projection is mainly to build linear proportionality among the relevant object-space disparities and with their corresponding elevations. This property allows generating supplementary disparity data to be used for filling the disparity gaps of the reference SDM. Additionally, the property allows the direct derivation of object-space elevations from the enriched (gap-free) SDM. These derived disparity-based elevations represent the LoS-DSM co-registered to the reference optical image selected. The DECR method includes analytical derivation of two formulas: (1) Disparity-to-

Elevation Formula (DEF) that provides the elevation values and (2) the Correct Distance Formula (CDF) that provides the shift to the correct geo-referenced location of the object-space pixels in the epipolar direction.

The developed DECR method was successfully validated in terms of the epipolarity condition of the projected images, the disparity proportionality among the projected images, the elevation accuracy derived based on the generated disparity data, and the elevation co-registration accuracy to one of the employed optical images (the selected reference image).

For applicability demonstration purposes, an EBD procedure was developed and implemented to detect building roofs using the achieved disparity-based LoS-DSM. The detection result for the off-terrain objects was achieved by a threshold value close to zero above the selected average terrain level. This represents an advantage over the original LoS-DSM algorithm. When the detection was evaluated, the result was found to be very successful based on the traditional performance measures (Completeness, Correctness, and Overall Quality). This indicates an effective minimization of the terrain-relief variation which reduces the need for an elevation normalization algorithm in dense urban areas of moderate terrain variation.

The Overall Quality measure was found to be 92% proving the successful accuracy of both disparity-based elevation data calculation and the optical-elevation data co-registration. While the 95% correct detection is attributed to the high quality of the elevation data generation and incorporation, the 96% complete detection resulted from the refined detection accuracy based on the occlusion and façades bitmap generated based on the Left-Right checking technique.

179

In addition to the high detection quality achieved, the generated building-roof objects were geo-referenced to their correct object-space location to allow direct integration with existing GIS layers. The geo-referencing process is based on moving the detected building roofs to the correct ortho-location in the epipolar domain of the reference image. The calculated locations were validated using an orthoimage generated for the study area using a commercial photogrammetric software package (PCI Geomatica). The evaluation result was almost 100% correct based on the visual assessment of the geo-referenced roof objects of which ≥90% overlap with the correct ground location of the corresponding building roofs in the orthoimage. This significantly successful geo-referencing result of the generated building roof map allows a direct integration with the existing layers of the GIS systems.

The identified limitation in the developed DECR method is its applicability to only the MVIS-VHR images acquired by push-broom linear sensors. Moreover, the terrain variation minimization by re-projection onto the average terrain level may not be sufficient in hilly and mountainous areas. For the developed EBD procedure, the identified limitation is the vegetation detection and removal based on NDVI-based threshold. This vegetation index may produce some errors in the cases of roof gardens and buildings with high NDVI values.

Therefore, future work should address these limitations and extend the DECR technique to be applicable in a tile-based approach in order to work even on hilly and large areas.

ACKNOWLEDGEMENTS

REFERENCES

Avbelj, J., Iwaszczuk, D., Müller, R., Reinartz, P., & Stilla, U. (2015). Coregistration refinement of hyperspectral images and DSM: An object-based approach using spectral information. *ISPRS Journal of Photogrammetry and Remote Sensing, 100*(2), 23-34.

Awrangjeb, M., Zhang, C., & Fraser, C. (2013). Automatic extraction of building roofs using LIDAR data and multispectral imagery. *ISPRS journal of photogrammetry and remote sensing, 83*(9), 1-18.

Awrangjeb, M., Ravanbakhsh, M., & Fraser, C. (2010). Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing, *65*(5), 457-467. doi: 10.1016/j.isprsjprs.2010.06.001

Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. And Griesbner,G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*, (pp. 12-23). Wichmann Verlag, Karlsruhe.

Dey, V. (2013). Image Segmentation Techniques for Urban Land Cover Segmentation of VHR Imagery: Recent Developments and Future Prospects. *International Journal of Geoinformatics*, *9*(4), 15-35.

Egnal, G., & Wildes, R. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(8), 1127-1133.

Garcia-Castellanos, D., & Lombardo, U. (2007). Poles of inaccessibility: A calculation algorithm for the remotest places on earth. *Scottish Geographical Journal, 123*(3), 227-233. doi: 10.1080/14702540801897809

Goshtasby, A. (2012). Image Registration Principles, Tools and Methods. Springer Science and Business Media. London

Guoqing, Z., Weirong, C., Kelmelis, J., & Deyan, Z. (2005). A comprehensive study on urban true orthorectification. *IEEE Transactions on Geoscience and Remote Sensing, 43*(9), 2138-2147. doi: 10.1109/TGRS.2005.848417

Habib, A., Ghanma, M., Morgan, M., & Al-Ruzouq, R. (2005). Photogrammetric and lidar data registration using linear features. *Photogrammetric Engineering and Remote Sensing, 71*(6), 699-707.

Habib, A., Kim, E., & Kim, C. (2007). New methodologies for true orthophoto generation. *Photogrammetric Engineering and Remote Sensing, 73*(1), 25-36.

Kwak, T., Kim, Y., Yu, K., & Lee, B. (2006). Registration of aerial imagery and aerial LiDAR data using centroids of plane roof surfaces as control information. *KSCE journal of Civil Engineering,* 10(5), 365-370.

Lowe, G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

Mishra, R., & Zhang, Y. (2012). A Review of optical imagery and airborne LiDAR data registration methods. *The Open Remote Sensing Journal, 5*, 54-63.

Parmehr, E., Fraser, C., Zhang, C., & Leach, J. (2014). Automatic registration of optical imagery with 3D LiDAR data using statistical similarity. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*(2), 28-40.

Rottensteiner, F., Trinder, J., Clode, S., & Kubik, K. (2005). Using the dempster–shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion, 6*(4), 283-300. doi: 10.1016/j.inffus.2004.06.004

Rutzinger, M., Rottensteiner, F., & Pfeifer, N. (2009). A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2*(1), 11-20.

Salehi, B, Zhang, Y, & Zhong, M. (2011). Object-based land cover classification of urban areas using VHR imagery and photogrammetrically-derived DSM. *Proceeding of the ASPRS Annual Conference*, Milwaukee, WI, USA.

Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of Geoinformatics, 8*(2), 35-51.

Story, M., & Congalton, R. (1986). Accuracy assessment–A user's perspective. *Photogrammetric Engineering and Remote Sensing, 52*(3), 397-399. doi: citeulike-article-id:4284683

Suliman, A., & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(5), 1913-1923.

Suliman, A., & Zhang, Y. (2016). Double projection planes method for generating enriched disparity maps from multi-view stereo satellite images. *The photogrammetric Record,*(Under Review).

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016). Registration-based mapping of aboveground disparities (rmad) for building detection in off-nadir vhr stereo satellite imagery. *Photogrammetric Engineering and Remote Sensing, 82*(7), 535-546.

Teillet, P. (1992). An algorithm for the radiometric and atmospheric correction of AVHRR data in the solar reflective channels. *Remote Sensing of Environment, 41*(2), 185-195.

Wong, A., & Orchard, J. (2008). Efficient FFT-accelerated approach to invariant optical–LIDAR registration. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(11), 3917-3925. doi: 10.1109/TGRS.2008.2001685

Xiong, Z., & Zhang, Y. (2010). A critical review of image registration methods. *International Journal of Image and Data Fusion, 1*(2), 137-158.

Zhang, J., Cao, Y., Zheng, Z.g, Chen, C., & Wang, Z. (2014). A new closed loop method of super-resolution for multi-view images. *Machine Vision and Applications, 25*(7), 1685-1695.

Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering and Remote Sensing, 70*(6), 657-661.

Zitová, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing, 21*(11), 977-1000. doi: 10.1016/S0262-8856(03)00137-9

Chapter 6: SUMMARY AND CONCLUSION

This chapter summarizes the research presented in this dissertation. It begins with the summary of each chapter (**Chapters 2 to 5**). The achievements of this research are then presented. Finally, recommendations for future work are provided.

6.1   Summary of Research

In this dissertation, stereo off-nadir VHR satellite images were exploited for 3D-based building detection using stereo information. Both the elevation and disparity information, which were extracted from stereo images, were utilized for detecting buildings in dense urban areas. **Chapters 2-5** introduced progressively improved methods addressing the four identified challenges associated with the use of off-nadir images. While elevation data were incorporated effectively in **Chapter 2** for elevation-based building detection, disparity data were successfully exploited in **Chapter 3** for disparity-based building detection. **Chapter 4** extended the concept of **Chapter 3** to utilize multi-view stereo VHR satellite images. Finally, the advantages of the methods developed in the three preceding chapters were combined and extended in **Chapter 5**. The progressive improvement is also demonstrated through the building detection results as provided in **Table 6.1.** The table shows improvements in the Overall Quality measure.

Table 6.1 Summary of the building detection results achieved in **Chapters 2-5**

Chapter No.	*Comp. (%)*	*Corr. (%)*	*OQ. (%)*
Chapter 2	89.0	77.5	70.7
Chapter 3	98.0	74.0	75.0
Chapter 4	-	-	-
Chapter 5	96.0	95.0	92.0

Chapter 2 introduced a solution for the problem of co-registering orthographic elevation data of digital surface models (DSM) with off-nadir VHR satellite images. This problem is caused mainly by the differences in the projections of the two data types being co-registered. While the DSMs are commonly in orthographic projections, off-nadir VHR images are acquired with perspective projections which create building lean.

The solution introduced in this chapter is to generate a line-of-sight DSM (LoS-DSM) that has the same perspective projection as the off-nadir image being overlaid in order to efficiently achieve accurate co-registration and preserve the original image information without implementing sophisticated algorithms (e.g., algorithms for occlusion detection and compensation) as in the case of true orthorectification.

The LoS-DSM was generated using the sensor model information of the image being co-registered in two major phases: (1) deriving DSM photogrammetrically from stereo images, and (2) re-projecting the derived DSM back to one of the stereo images for accurate co-registration. The achieved dataset after this image-elevation co-registration forms the bases for mapping off-terrain features even in off-nadir VHR images. Subsequently, the applicability of the developed LoS-DSM was demonstrated through elevation-based building detection in off-nadir VHR satellite imagery.

With the aim of decreasing the computational steps required for detecting buildings, **Chapter 3** replaces photogrammetric-based elevation data with disparity information to reduce the computational steps and costs. Moreover, since the generated disparity maps fit exactly the selected reference image, no effort was required for image-disparity co-registration. However, given that disparity information is typically mapped from a pair of stereo images, a stereo pair of off-nadir VHR images over dense urban areas suffers from

a lot of occlusion areas. Consequently, many occlusion disparity gaps will be created in the generated disparity maps. Interpolating the resulting gaps and normalizing the surface disparity information produces misleading surface information that destroys the subsequent processes including building detection.

Therefore, this chapter introduced a registration-based method for mapping the disparities of the aboveground features. The method attempts to bypass the steps of interpolating the gaps and normalizing the surface information by co-registering the corresponding terrain-level features in an epipolar stereo pair in order to minimize the disparity values of the terrain-level features. By this registration-based minimization, the mapped disparities from the constructed epipolar stereo pair will be close to zero disparity values for the terrain features. Hence, only the disparities of the elevated/ aboveground objects will be measured by all epipolar-based matching techniques. The achieved dataset of the co-registered and normalized disparity map form the bases for mapping off-terrain features in off-nadir VHR stereo images acquired over dense and reasonably non-flat urban areas. Therefore, the applicability of the developed registration-based method was demonstrated through a disparity-based building detection in off-nadir VHR satellite imagery.

Chapter 4 represents an extension to the last introduced solution for disparity-based building detection. In this chapter, the two common matching problems of occlusion producing gaps and confusion producing outliers were addressed by generating supplementary disparity data from multi-view inline stereo off-nadir satellite images to allow disparity map enrichment (i.e., filling the disparity gaps). The generation of the

186

supplementary data was made possible by utilizing the two models of the sensor information of each stereo image employed for analytically building inter-proportionality in the object space among the disparities from different stereo pairs. This proportionality allowed supplementary disparity data to be generated. Hence, an enriched disparity map was successfully achieved

Building on the methods presented in the preceding chapters, **Chapter 5** introduced an extended and improved combined approach. The concept developed in **Chapter 4** was exploited to derive elevation information from disparity data that are co-registered to the selected reference image. Compared to the solution developed in **Chapter 2**, **Chapter 5** introduced a further improved method based on disparity information to generate LoS-DSM for image-elevation co-registration. The applicability of the disparity-based generation of the LoS-DSM was demonstrated through elevation-based building detection in off-nadir VHR satellite imagery.

Since the building detection maps in off-nadir VHR satellite images cannot be geo-referenced directly due to the building-roof offsets from their corresponding footprints in off-nadir images, **Chapter 5** derived a formula based on the disparity-based generated elevations for correcting these offsets. Based on this formula, the rooftop offsets were corrected and the detection map was geo-referenced successfully using a conventional registration based on a first order polynomial transformation.

6.2   Achievements of the Research

A summary of the introduced technologies and achievements in each of the four main chapters is presented in the following subsections:

6.2.1   Developing the Line-of-Sight DSM Solution for Efficient Image-Elevation Co-registration (LoS-DSM Solution)

In **Chapter 2**, an efficient elevation co-registration with off-nadir VHR satellite imagery was successfully achieved to support building detection based on the line-of-sight digital surface model (LoS-DSM) solution. In the current literature, true orthorectification is the most successful method that can be used to co-register image and elevation data. However, true orthorectification requires accurate DSM data, occlusion detection and compensation algorithms, and multiple overlapped images to disclose the hidden areas. In comparison to the true orthorectification process, the developed LoS-DSM solution for image-elevation co-registration is more efficient and has a few advantages. The solution does not require (1) high quality DSM data, (2) sophisticated algorithms, or (3) multiple stereo pairs. Moreover, the solution (4) provides sub-pixel co-registration accuracy level, (5) preserves the original image information, and (6) is applicable regardless of the acquisition angle of the images being co-registered. In addition to these advantages, the solution is (7) highly compatible with the elevation-based building detection applications. An improvement in building detection quality of almost 12% was achieved in comparison to the same detection procedure when the conventional 2D image registration method was used with the off-nadir images acquired even over high-rise building areas.

6.2.2 Introducing the Image Registration Concept for Mapping Aboveground Disparities (RMAD Method)

In **Chapter 3**, a more computationally cost-effective method, than that presented in the previous chapter, for building detection in off-nadir VHR satellite image was successfully achieved based on disparity data. Replacing photogrammetrically-derived elevations by disparity measurements saves most of the typical photogrammetric processes as well as the co-registration effort since the generated map fits exactly the reference image. The developed method mitigates the severe occlusion effects in off-nadir VHR satellite stereo images acquired over reasonably non-flat urban areas by attempting to directly measure the disparities of the elevated objects. Such measurements are achieved after minimizing the terrain disparities based on registering the terrain-level features in both off-nadir VHR images of the employed epipolar stereo pair.

The main advantage achieved by this method is the efficient reduction of implementation costs required for building detection through bypassing the steps of (1) stereo information co-registration, (2) occlusion-produced gap interpolation, and (3) surface disparity normalization. Additionally, (4) the method reduces the required input image data to only one stereo pair of off-nadir VHR satellite images. Finally, (5) the developed method is highly compatible with disparity-based building detection applications.

For dense urban areas with many high-rise buildings, the achieved quality improvement of building detection in off-nadir VHR satellite images based on the developed method was more than 30% in comparison to the same detection procedure when the typical interpolation and normalization techniques were executed.

6.2.3 Developing the Double Projection Planes Method for Supplementary Disparity Data Generation (DPP Method)

In **Chapter 4**, disparity proportionality among different stereo pairs of multi-view VHR stereo satellite image was achieved. This proportionality was constructed using the two forms of the sensor model information in each stereo image employed to allow calculating disparity transferring scales and hence allow supplementary disparity data to be generated. These supplementary data allowed filling the disparity gaps of the measured maps in dense urban areas, and hence, the negative effects of executing interpolation techniques were minimized. Additionally, the supplementary data allowed filling the gaps of the detected and removed outliers, thereby avoiding the drawbacks of designing and using convolution filters to both filling gaps and correcting outliers. Accordingly, a gap-and-outlier-free disparity map was achieved which significantly assists the applications of disparity-based building detection.

6.2.4 Developing the Disparity-based Method for Generating the Line-of-Sight DSM Solution (DECR Method)

In **Chapter 5**, a more complete and improved generation for the line-of-sight digital surface model (LoS-DSM) solution developed in **Chapter 2** was achieved. The improved solution reduced the computations required in the original LoS-DSM algorithm of **Chapter 2** and at the same time preserved the pixel-level accuracy of image-elevation co-registration. Additionally, the improved solution extended the achievement of **Chapter 4** by deriving elevation data from the enriched disparity map generated from multi-view inline off-nadir VHR satellite images (this feature is not available in the

method of **Chapter 3**). This was achieved by constructing a proportional relationship between ground elevations and their corresponding disparities which allowed disparity-based elevations to be generated.

Due to the quality of the generated elevations based on the enhanced disparity map, a high overall quality, of more than 90%, of elevation-based building detection was achieved in an off-nadir VHR satellite image acquired over a challenging urban area. To allow direct geo-referencing for the detection maps, a formula to correct the offsets of the building roofs mapped in off-nadir VHR images was derived. By applying this offset correction, almost 100% of the mapped roofs were geo-referenced correctly.

Based on these achievements, the detection maps should be easily integrated with the existing layers of the GIS systems and hence, the usability and benefits of off-nadir VHR satellite images would be significantly increased. This will expand the data source scope for building detection application to include the off-nadir VHR images archived in the online databases of the satellite image vendors.

6.3 Limitations and Recommendations for Future Work

Based on the results and contributions discussed in the previous sections, the identified limitations and suggested recommendations for future research are as follows:

- In **Chapter 2**, the introduced LoS-DSM solution for co-registering an off-nadir VHR satellite image with elevation data is developed based on DSM elevations derived photogrammetrically from a stereo pair, one of whose images is used for the co-registration. In the case of DSM data that are pre-existing from other sources (e.g., LiDAR-based DSM), the method needs to be modified to

manipulate the occlusion problem in the pre-existing DSM. For example, the occluded areas can be addressed by selecting the elevation values that are visible by the satellite sensor identified by visibility checks. Hence, a suggested future work is to overcome the challenges associated with employing pre-existing DSM data in the co-registration and building detection procedures.

- In **Chapter 3**, the introduced registration-based method for mapping the disparities of the aboveground objects is limited to reasonable terrain variations. In the case of hilly areas, defining the projection reference plane by the average elevation for the area will not be enough to eliminate the terrain disparities. Hence, the suggested future work is to work with small areas by adopting a tile-based approach using the online freely available elevation data to co-register terrain-level features for implementing the developed registration-based method.

- In **Chapter 4**, the introduced double projection planes method is developed for VHR images acquired by linear sensors. If the image data were acquired by a frame sensor, the developed method would encounter some difficulties such as losing the epipolarity condition in the object space if the images were acquired with different and large tilt angles. Additionally, the method was tested using multi-view inline stereo VHR satellite images. Therefore, the suggested future work is to address the challenges of applying the method and extending the algorithm to accommodate frame sensors by mainly addressing the problem of the scale changes for the tilted images. Furthermore, the method may be tested using across-track stereo VHR satellite images to identify and address any associated challenges.

192

- In **Chapter 5**, the introduced disparity-based elevation generation and co-registration method is developed for VHR images acquired by linear sensors over dense urban areas of reasonable terrain variations. Hence, the same limitations identified in Chapter 4 apply. Additionally, the method is developed to derive elevations based on a scale-relationship with the corresponding disparity data. The relationship requires further investigation and testing to identify its limitations. Therefore, the suggested future work is to further investigate disparity-to-elevation relationship in a variety of cases and scenarios. Moreover, the future work should also investigate the possibility of extending the algorithm to accommodate frame sensors and adopt the tile-based approach to mitigate the problems of hilly areas. Furthermore, the tests should include applying the developed method to across-track stereo VHR satellite images to identify and address the encountered challenges.

- In three chapters (**Chapter 2, 3, and 5**), the building detection procedure includes a vegetation detection and removal process since the elevated trees may be wrongly detected as building objects. The detection of the vegetation is based on calculating and thresholding a spectral-based vegetation index (i.e., NDVI). This vegetation index may produce some errors in the cases of roof gardens and buildings with high NDVI values. Hence, a suggested future task is to identify a more reliable detection feature for vegetation, such as vegetation texture information, to overcome this limitation in the implemented building detection procedure.

APPENDIX I

In this appendix, the information regarding implementing the segmentation step followed in this dissertation is provided. The segmentation algorithm executed in building detection process in the dissertation **Chapters 2, 3 and 5** is the multi-resolution segmentation (MRS) as introduced by Baatz and Schäpe (2000)[1]. This algorithm is provided within the environment of the *eCognition®* software package. The package version used for the presented research work is v 9.0.

The MRS algorithm requires three parameters: *Scale*, *Shape*, and *Compactness*. The segmentation results rely directly on the values selected for these parameters. The common way to find these values is by following the trial-and-error technique until we get the most acceptable result. However, this technique is very time consuming. Hence, the followed technique in this dissertation is to over-segment the VHR image employed for building detection using a small scale value (≈ 60) and the default values for the other two parameters (Shape = Compactness = 0.5). By over-segmenting the image, the edges of the building rooftops will be preserved. After that, the elevated objects are thresholded. The resulting detection needs to be finished and enhanced by applying the following steps:

1. Detecting and removing the tree objects by using NDVI vegetation index.

---

[1] Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. And Griesbner,G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*, (pp. 12-23). Wichmann Verlag, Karlsruhe.

2. Detecting and removing the building sides by using the Left-Right visibility checking technique to generate an occlusion bitmap mask.

3. Merging the detected building objects.

4. Filling the areas that are surrounded by building objects.

5. Removing the small objects that usually represent noise by applying an empirically-selected threshold value to detect and exclude these segments based on their areas.

6. Finally, applying morphological operations to finish and enhance the detection results such as smoothing the boundaries of the detected building objects. A variety of enhancement operations and algorithms based on the Cognition network Language (CNL) are available within the environment of the *eCognition®* 9.0 software package.

APPENDIX II

Permission from IEEE Journal of Selected Topics in Applied Earth Observation and

Remote Sensing (IEEE-J-STARS):

Dear Alaeldin Suliman,

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant: *Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:*

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:
1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to: http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

APPENDIX III

Permission from Photogrammetric Engineering and Remote Sensing (PE&RS):

A permission form the journal publisher was granted by Dr. Michael Hauck on May 2nd, 2016 to use the following paper in this PhD dissertation:

Suliman, A., Zhang, Y., & Al-Tahir, R. (2016). Registration-based Mapping of Aboveground Disparities (RMAD) for Building Detection in Off-nadir VHR Stereo Satellite Imagery. *Photogrammetric Engineering and Remote Sensing*, 82(7): 535-546.

The received email is provided below:

Dear Mr. Suliman,

   Congratulations on nearing the completion of your doctoral studies. ASPRS grants permission as you requested below.
Best wishes,
Michael

Dr. Michael Hauck
Executive Director, ASPRS
Publisher, PE&RS

www.asprs.org

Curriculum Vitae

Candidate's full name: Alaeldin Suliman

Universities attended (with dates and degrees obtained):

2014: Diploma in University Teaching, University of New Brunswick, Canada

2007: M.Sc., Civil-Surveying Engineering, University of Benghazi, Libya

2001: B.Sc., Civil Engineering, University of Garyounis, Benghazi, Libya

Publications:

Peer-Reviewed Journal Papers

1. **Suliman, A.**, & Zhang, Y. (2015). Back-propagation neural networks in remote sensing image classification – A review on the design and implementation of the networks. *Journal of Earth Science and Engineering*, *5*(1), 52-65.
2. **Suliman, A.**, & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(5), 1913-1923.
3. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). Registration-based mapping of aboveground disparities (RMAD) for building detection in off-nadir VHR stereo satellite imagery. *Photogrammetric Engineering & Remote Sensing*, *82*(7), 535-546.
4. **Suliman, A.**, & Zhang, Y. (2016). Segment-based terrain filtering technique for elevation-based building detection in VHR remote sensing images. *Advances in Remote Sensing*, *5*(3), 2016.
5. Sharaf-El-Din, E., Zhang, Y., & **Suliman, A**. (2016). Mapping the concentrations of surface water quality parameters using a remote sensing and artificial intelligence framework. *International Journal of Remote Sensing*. **In Press**.
6. **Suliman, A.**, & Zhang, Y. (2016). Double projection planes method for generating enriched disparity maps from multi-view stereo satellite images. *The photogrammetric Record*, (Under Review).
7. **Suliman, A.**, & Zhang, Y. (2016). Disparity-based generation of line-of-sight DSM for image-elevation co-registration to support building detection in off-nadir VHR satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, (Under Review).
8. **Suliman, A.**, & Zhang, Y. (2016). Optical-elevation data co-registration and classification-based height normalization for building detection in stereo VHR images. *GEOMATICA*. (Under Review).

Conference Papers

1. **Suliman, A.**, & Zhang, Y. (2014). Integration of off-nadir VHR imagery with elevation data for advanced information extraction. *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 13-18 July, Quebec, Canada.
2. Jabari, S., Zhang, Y., & **Suliman, A**. (2014). Stereo-based building detection in very high resolution satellite imagery using IHS color system. *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 13-18 July, Quebec, Canada.
3. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). A novel technique for mapping the disparity of off-terrain objects. *IOP Conference Series: Earth and Environmental Science, 34*(1), 012035.
4. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). Co-registering and normalizing stereo-based elevation data to support building detection in VHR images. *Proceedings of the Imaging & Geospatial Technology Forum (IGTF). The ASPRS annual conference 2016*. April 11-15, Fort Worth, US.
5. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). Slope-based terrain filtering for building detection in remotely sensed VHR images. *Proceedings of the Imaging & Geospatial Technology Forum (IGTF). The ASPRS annual conference 2016*. April 11-15, Fort Worth, US.
6. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). Enhanced disparity maps from multi-view satellite images. *Proceedings of the 2016 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*. July 10-15, Beijing, China.
7. **Suliman, A.**, Zhang, Y., & Al-Tahir, R. (2016). Extracting accurate building information from off-nadir VHR images. *Proceedings of the Ninth International Conference on Geographic Information Science (GIScience)*. September 27-30, Montreal, Canada.
8. **Suliman, A.**, & Zhang, Y. (2017). An efficient approach for image-DSM co-registration for urban building extraction. *Proceedings of the Biennial International Joint Urban Remote Sensing Event (JURSE)*, 5-7 March, Dubai, United Arab Emirates (*Accepted on November 18/2016*).

Conference Presentations:

1. **Suliman, A.**, & Zhang, Y. "*A polygon-based filtration technique for building objects' detection*", The 2013 Geomatics Atlantics conference, September 23-25, Saint John, Canada.
2. **Suliman, A.**, & Zhang, Y. "*Efficient registration method for integrating elevation data with off-nadir satellite imagery*", The 2014 Spring Graduate Seminar and Student technical Conference, Geodesy and Geomatics department, UNB, April 15, Fredericton, Canada.
3. **Suliman, A.**, & Zhang, Y. "*Elevation-based building detection strategy*", the 2014 Graduate Research conference (GRC), UNB, April 24, Fredericton, Canada.
4. **Suliman, A.**, & Zhang, Y. "*A novel approach for measuring the disparity information of aboveground objects to support building detection in off-nadir VHR*

satellite images", The 2015 Spring Graduate Seminar & Student technical Conference, Geodesy and Geomatics department, UNB, March 25, Fredericton, Canada.

5. **Suliman, A.**, & Zhang, Y. "*A novel disparity-based building detection*", The 2015 Graduate Research conference (GRC), UNB, April 23, Fredericton, Canada. (***Award winning presentation***)

6. **Suliman, A.**, & Zhang, Y. "*An integrated technique for supporting stereo-based building detection in VHR remote sensing images*", The 2016 Graduate Research conference (GRC), UNB, April 22, Fredericton, Canada.

7. **Suliman, A.**, & Zhang, Y. "*Object-space disparity concept for enriched disparity maps from multi-view satellite images*", The 2016 Graduate Research conference (GRC), UNB, April 22, Fredericton, Canada. (***Award winning presentation***)

Chairing Technical Sessions:

- **Science and Engineering** at the 2016 Graduate Research conference (GRC), UNB, April 22, Fredericton, Canada.