

MATCHING AND RECONSTRUCTION OF LINE FEATURES FROM  
ULTRA-HIGH RESOLUTION STEREO AERIAL IMAGERY

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

BY

ALİ ÖZGÜN OK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN  
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

SEPTEMBER 2011

Approval of the thesis:

**MATCHING AND RECONSTRUCTION OF LINE FEATURES FROM  
ULTRA-HIGH RESOLUTION STEREO AERIAL IMAGERY**

submitted by **ALİ ÖZGÜN OK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen  
Dean, Graduate School of **Natural and Applied Sciences** \_\_\_\_\_

Assoc. Prof. Dr. Ayşegül Aksoy  
Head of Department, **Geodetic and Geographic Information Techn.** \_\_\_\_\_

Prof. Dr. Vedat Toprak  
Supervisor, **Geodetic and Geographic Information Techn. Dept., METU** \_\_\_\_\_

Prof. Dr. Uwe Soergel  
Co-Supervisor, **Inst. of Photogrammetry and Geoinformation, Hannover Univ.** \_\_\_\_\_

**Examining Committee Members:**

Prof. Dr. Volkan Atalay  
Computer Engineering Dept., METU \_\_\_\_\_

Prof. Dr. Vedat Toprak  
Geological Engineering Dept., METU \_\_\_\_\_

Assoc. Prof. Dr. Mustafa Türker  
Geomatics Engineering, HÜ \_\_\_\_\_

Dr. Franz Rottensteiner  
Inst. of Photogrammetry and Geoinformation, Hannover Univ., Germany \_\_\_\_\_

Dr. Uğur Murat Leloğlu  
Institute Head, TUBITAK Space Technologies Research Institute \_\_\_\_\_

**Date:** 13. 09. 2011

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

Name, Last name : Ali Özgün Ok

Signature :

## **ABSTRACT**

### **MATCHING AND RECONSTRUCTION OF LINE FEATURES FROM ULTRA-HIGH RESOLUTION STEREO AERIAL IMAGERY**

OK, Ali Özgün

PhD., Department of Geodetic and Geographic Information Technologies

Supervisor: Prof. Dr. Vedat TOPRAK

Co-Supervisor: Prof. Dr. Uwe SOERGEL

September 2011, 196 pages

In this study, a new approach for the matching and reconstruction of line features from multispectral stereo aerial images is presented. The advantages of the existing multispectral information in aerial images are fully taken into account all over the steps of pre-processing and edge detection. To accurately describe the straight line segments, a principal component analysis technique is adapted. The initial correspondences between the stereo images are generated using a new pairwise stereo matching approach which involves a total of seven relational constraints. The final line to line correspondences between the stereo images are established in a precise matching stage in which the final line matches are assigned by means of three novel measures and a final similarity voting scheme. Once the line matches are established, the stereo reconstruction of those matches is performed by an innovative reconstruction approach that manipulates the redundancy inherent in line pair-relations. By this way, the reconstruction of the stereo matches that are observed in a nearly-parallel geometry with the epipolar

lines can also be performed accurately.

The proposed approach is tested over two different urban test sites with various built-up characteristics, and as a result, very successful and promising stereo line matching and reconstruction performances are reached. Besides, the comparison of the results of the proposed approach with the results of one of the state-of-the-art stereo matching approaches proves the superiority and the potential of proposed approach.

Keywords: Pair-wise Line Matching, Reconstruction, Straight Line Extraction, Stereo Aerial Images, Photogrammetry.

## ÖZ

### ULTRA YÜKSEK ÇÖZÜNÜRLÜKLÜ STEREO HAVA FOTOĞRAFLARINDAN DOĞRUSAL NESNELERİN EŞLENMESİ VE GERİ ÇATIMI

OK, Ali Özgün

Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü

Tez Yöneticisi: Prof. Dr. Vedat TOPRAK

Ortak Tez Yöneticisi: Prof. Dr. Uwe SOERGEL

Eylül 2011, 196 sayfa

Bu çalışmada, çok-bantlı stereo hava görüntülerinden elde edilen çizgisel nesnelere eşlenmesi ve üç boyutlu geri-çatımı için yeni bir yaklaşım sunulmuştur. Hava görüntülerinde var olan çok-bant bilgisi, ön-işleme ve kenar bulma aşamalarında tam olarak kullanılmıştır. Doğrusal çizgilerin başarılı bir şekilde çıkarılabilmesi amacıyla temel bileşenler analizi yöntemi uygulanmıştır. Stereo görüntülerdeki ilk eşlemeler, toplamda yedi adet ilişkisel koşul içeren yeni bir çift-tabanlı stereo çizgi eşleme yöntemi ile gerçekleştirilmiştir. Stereo görüntülerdeki nihai bire-bir çizgi eşlemeleri, üç adet yeni ölçüt ve son-benzerlik değerlendirilmesi kısımlarını içeren hassas eşleme aşamasında belirlenmiştir. Çizgi eşlemeleri bulunduktan sonra bu çizgilerin üç boyutlu geri-çatımı, çift-tabanlı çizgisel ilişkilerin doğasında bulunan ve mükerrer eşleme bilgisini kullanan yenilikçi bir yaklaşım ile gerçekleştirilmiştir. Bu sayede, epipolar çizgisine neredeyse paralel olarak bulunan stereo eşlemelerin geri-çatımının da başarılı bir şekilde yapılabilmesi mümkün olabilmektedir.

Önerilen yaklaşım çeşitli yerleşim şekilleri barındıran iki farklı test alanında değerlendirilmiş ve oldukça başarılı ve umut verici stereo çizgi eşleme ve geri-çatım sonuçlarına ulaşılmıştır. Bunun yanı sıra, önerilen yaklaşımın sonuçları ile çizgi eşlemede başarısı kabul edilmiş ve bilinen bir yöntemin sonuçlarının karşılaştırması da önerilen yaklaşımın üstünlüğünü ve potansiyelini kanıtlamaktadır.

Anahtar Kelimeler: Çift-tabanlı Çizgi Eşleme, Geri-çatım, Doğrusal Çizgi Çıkarımı, Stereo Hava Görüntüleri, Fotogrametri.

To my Parents

and

my Wife

## ACKNOWLEDGMENTS

I would like to gratefully thank to my supervisor Prof. Dr. Vedat Toprak for his support and motivation during the development of this thesis and I wish him all the best for his new career move.

I wish to express my deepest gratitude to my co-supervisor Prof. Dr.-Ing. Uwe Soergel for his endless support, motivation and guidance. I will never forget your helps throughout this research and I would like to cordially thank you for your permanently positive and warm attitude especially during the six months period I spent in Hannover - Institute of Photogrammetry and GeoInformation (IPI).

I also cordially would like to thank Dr. Franz Rottensteiner for his valuable suggestions and contributions in this thesis. I'm also thankful to him for developing the code that we used during the evaluation of the 3D line segments with the LIDAR data. His very detailed review and comments to the written text considerably improved this thesis.

I am very grateful to Prof. Dr. Volkan Atalay and Assoc. Prof. Dr. Mustafa Türker for their valuable comments during the thesis supervising committees. I also would like to thank Dr. Uğur Murat Leloğlu for his useful suggestions to the written text.

I would like to send many thanks to Dr.-Ing. Jan Dirk Wegner and Prof. Dr.-Ing. Christian Heipke for their suggestions, encouragement and fruitful discussions we had particularly during my visit in Hannover. I also would like to thank Dr.-Ing. Karsten Jacobsen for his helps with the program BLUH.

I would like to thank and acknowledge the providers of the data sets used in this thesis. The Hannover data set was provided by the University of Hannover - Institute of Photogrammetry and GeoInformation (IPI). The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF).

I would like to thank to all my professors and instructors at the department of GGIT. I also would like to thank all my friends in GGIT, Dr. Kıvanç Ertuğay, Dr. Serkan Kemeç, Dr. Dilek Koç San, Özgün Balkanay, Dr. Resat Geçen, Pınar Aslantaş Bostan, Dr. Arzu Erener, Gülcan Sarp, Serdar Sürer, Önder Gülbeyaz and Nuhcan Akçit for their friendship, helps and encouragement during the time I spent in METU.

I am thankful to all of the friends I met at IPI. My special thanks go to Dr.-Ing. Jan Dirk Wegner and Dipl.-Ing. Joachim Niemeyer for their close friendship and continuously brother-like attitude. Once again, I appreciate all the help of my roommate, Dipl.-Ing. Joachim Niemeyer, and I have to admit that without his help and friendship, the life at Hannover would be much more difficult.

I wish to express my deepest gratitude to Prof. Dr. Ülkü Yetiş and Prof. Dr. Filiz B. Akın for their endless support and motivation. I want you to know that without your positive approach, this research would be much more challenging.

I would like to thank Cem Güllüoğlu, İlksen Urgancı Güllüoğlu, Assist. Prof. Dr. Emre Sümer for their motivation and close friendship during this study. I also would like to appreciate the helps and information given by Assist. Prof. Dr. Umut Güneş Sefercik for my Hannover research visit. Special thanks go to M. Alper Temel for providing the finest wines for the celebration party of this thesis.

I would like to express my deepest gratitude to Sadi and İhsan Yetiş, my maternal uncles, and special thanks go to Tuğçe, Serra, Begüm and Mediha Yetiş for supporting me all the time. They do really mean a lot to me and thanks for everything.

I cordially would like to dedicate this thesis to my dear parents Nuri Ok and Semra Ok. Their continuous affection, encouragement, patience and belief made me who I am. I am really glad to have a father and mother like you. Thank you for everything. I also would like to thank my father-in-law Akil Özdarıcı and mother-in-law Fatma Özdarıcı for their encouragement and support during the preparation of this thesis. I am also thankful to my sisters-in-law Öznur Özdarıcı and Gül Özdarıcı for their supportive attitude and helps especially for the celebration party.

Last but not least, I would like to express my special thanks to my wife Aslı Özdarıcı Ok for her love, encouragement, help and understanding. Her supportive attitude gave me strength to overcome the difficulties I faced throughout this study, and I am really lucky to have a wife like you. Thank you for everything.

## TABLE OF CONTENTS

ABSTRACT .....	iv
ÖZ .....	vi
ACKNOWLEDGEMENTS .....	ix
TABLE OF CONTENTS .....	xii
LIST OF TABLES .....	xv
LIST OF FIGURES .....	xvi

### CHAPTERS

1. INTRODUCTION.....	1
1.1 Purpose and Scope .....	1
1.2 Contributions.....	4
1.3 Organization of the Thesis .....	6
2. BASICS AND STATE-OF-THE-ART .....	7
2.1 Optical Airborne Cameras.....	7
2.2 Edge and Line Matching .....	13
2.2.1 Related Work .....	14
2.2.2 Discussion of the Related Work.....	28
2.3 A Short Comparison of the Vision Techniques Utilized .....	30
3. PRE-PROCESSING AND 2D LINE SEGMENT EXTRACTION.....	36
3.1 Pre-Processing.....	38
3.1.1 Multi-Level Non-Linear Color Diffusion Filter.....	38
3.1.2 Color Boosting .....	44
3.2 Line Extraction.....	48
3.2.1 Color Edge Detection.....	48
3.2.2 Straight Line Extraction .....	51
4. STEREO LINE MATCHING .....	67

4.1	Formation of Potential Matching Candidates on the Base Image.....	68
4.2	Identification of Candidate Pair Models on the Search Image .....	72
4.3	Initial Pair-wise Matching.....	75
4.3.1	Geometric Constraints.....	75
4.3.2	Radiometric Constraint .....	78
4.3.3	Correlation Constraint.....	79
4.3.4	Regional Constraint.....	82
4.3.5	Epipolar Constraint .....	84
4.3.6	Pair-wise Matching .....	86
4.4	Precise Line-to-Line Matching .....	86
4.4.1	Line-based Daisy Measure .....	88
4.4.2	Redundancy Measure .....	94
4.4.3	Pair-wise Quality Measure .....	98
4.4.4	Precise Matching.....	98
5.	STEREO LINE RECONSTRUCTION.....	102
5.1	Reconstruction of Line Segments .....	103
5.1.1	The Generation of Uncertain Artificial 3D Point Entities .....	105
5.1.2	The Joint Estimation of the 3D Line Segments .....	107
6.	DISCUSSION: TEST DATA AND PERFORMANCE	
	EVALUATION.....	112
6.1	Test Sites and Image Datasets.....	112
6.1.1	The Vaihingen/Enz Test Site .....	112
6.1.2	The Hannover Test Site.....	114
6.2	Assessment Strategies for the Line Matching and Reconstruction....	115
6.2.1	Assessment Strategy for the Line Matching .....	115
6.2.2	Assessment Strategy for the Line Reconstruction .....	116
6.3	The Results and Discussion of the Proposed Line Matching	
	Approach.....	117
6.3.1	The Results and Discussion of the Vaihingen Test Site .....	117
	6.3.1.1 Parameter Selection for the Initial Pair-wise Matching	
	Stage.....	117

6.3.1.2	Parameter Selection for the Precise Matching Stage .....	123
6.3.1.3	Evaluation and Discussion of the Line Matching Results .....	129
6.3.2	The Results and Discussion of the Hannover Test Site .....	139
6.3.2.1	Parameter Selection for the Initial Pair-wise Matching Stage .....	139
6.3.2.2	Parameter Selection for the Precise Matching Stage .....	142
6.3.2.3	Evaluation and Discussion of the Line Matching Results .....	144
6.3.3	Comparison to the State-of-the-Art.....	150
6.4	The Results and Discussion of the Proposed Line Reconstruction Approach .....	156
7.	CONCLUSIONS AND RECOMMENDATIONS .....	167
7.1	Conclusions .....	167
7.2	Recommendations .....	171
	REFERENCES.....	173
	APPENDIX	
	A. Gauss-Helmert Model with Constraints.....	191
	CURRICULUM VITAE .....	194

## LIST OF TABLES

### TABLES

Table 2.1	Summary of the large format digital frame cameras.....	10
Table 2.2	Summary of the previous work on the topic of line matching using aerial and/or satellite images.....	15
Table 4.1	The results of the (a) pair-wise matching and (b) the inferred one-to-one line matches from the pair relations. ....	95
Table 6.1	Selected thresholds of the pair-wise constraints for the Vaihingen site. ....	122
Table 6.2	The total number of pair-wise relations observed before and after the initial pair-wise matching.....	123
Table 6.3	The selected thresholds for the precise matching of the Vaihingen site. ....	129
Table 6.4	Comparative matching results of the proposed approach and the Schmid's method for the Vaihingen test patches.....	152
Table 6.5	Comparative matching results of the proposed approach and the Schmid's method for the Hannover test patches. ....	154
Table 6.6	The computed RMS distances for the method of Direct Construction .....	159
Table 6.7	The computed RMS distances for the proposed approach.....	159

## LIST OF FIGURES

### FIGURES

Figure 1.1	Flowchart of the proposed approach .....3
Figure 2.1	Multiple camera image acquisition techniques .....11
Figure 3.1	Flowchart of the proposed 2D line extraction approach .....38
Figure 3.2	(a) Raw aerial image #1 and (b) the filtered result. (c-f) illustrates the evolution of the diffusion filtering process. (c) Original patch, (d) result of the first-level diffusion ( $\lambda = 0.03$ , $\sigma = 2$ ), (e) result of the second-level diffusion ( $\lambda = 0.05$ , $\sigma = 1$ ), and (f) result of the third-level diffusion ( $\lambda = 0.075$ , $\sigma = 0.5$ ). .....42
Figure 3.3	(a) Raw aerial image #2 and (b) the filtered result. (c-f) illustrates the evolution of the diffusion filtering process. (c) Original patch, (d) result of the first-level diffusion ( $\lambda = 0.03$ , $\sigma = 2$ ), (e) result of the second-level diffusion ( $\lambda = 0.05$ , $\sigma = 1$ ), and (f) result of the third-level diffusion ( $\lambda = 0.075$ , $\sigma = 0.5$ ). .....43
Figure 3.4	(a) Image #1 after discontinuity preserving filtering, and the results of color boosting with (b) $k = 2$ , (c) $k = 3$ , (d) $k = 5$ , (e) $k = 20$ , and (f) $k = 200$ . .....45
Figure 3.5	(a) Image #2 after discontinuity preserving filtering, and the results of color boosting with (b) $k = 2$ , (c) $k = 3$ , (d) $k = 5$ , (e) $k = 20$ , and (f) $k = 200$ . .....46
Figure 3.6	(a, d) The original images, boosted ( $k = 200$ ) images (b, e) after discontinuity preserving filtering, and (c, f) without performing any filtering operation.....47
Figure 3.7	(a) Original image in a grid pattern, (b) result of a color variant canny edge detector, and (c) result of a grayscale canny edge detector.....49

Figure 3.8	(a, c) Color boosted images with ( $k = 3$ ), and the results of color Canny edge detection with $\sigma = 2$ and hysteresis thresholds of $\tau_1 = 20$ and $\tau_2 = 0$ .....	52
Figure 3.9	(a) Artificial critical locations and (b) the corrected result.....	55
Figure 3.10	Masks utilized to identify and correct the artificial critical locations .....	55
Figure 3.11	(a) A problematic same-label intersection and (b) the corrected result.....	56
Figure 3.12	Masks utilized to identify and correct the critical intersections.....	56
Figure 3.13	(a) Original image #1, (b) filtered and boosted image, and (c) color Canny edges. Straight row, straight column and final straight edges are illustrated in first, second and third columns, respectively. (d-f) Straight edges with absolute thresholding with $t_a = 0.1$ , (g-i) straight edges with relative thresholding with $t_a = 0.1$ , and (j-m) straight edges with relative thresholding with $t_a = 1.6$ . For all cases, minimum segment length is selected as 10 pixels. ....	58
Figure 3.14	(a) Original image #2, (b) filtered and boosted image, and (c) color Canny edges. Straight row, straight column and final straight edges are illustrated in first, second and third columns, respectively. (d-f) Straight edges with absolute thresholding with $t_a = 0.1$ , (g-i) straight edges with relative thresholding with $t_a = 0.1$ , and (j-m) straight edges with relative thresholding with $t_a = 1.6$ . For all cases, minimum segment length is selected as 10 pixels. ....	59
Figure 3.15	Fitting a straight line to a set of edge pixels.....	60
Figure 3.16	Original images (a, d), straight line segments overlaid to straight edge segments (b, e), and straight line segments overlaid to original images (c, f).....	63
Figure 3.17	(a) An edge segment and the left and right flanking regions. (b, c) The generated flanking regions for each line segment.....	64
Figure 4.1	Flowchart of the proposed stereo line matching approach.....	67

Figure 4.2	The proximity measure for grouping the line pairs on the base image.....	69
Figure 4.3	The orientation measure for grouping the line pairs on the base image.....	70
Figure 4.4	The flanking regions measure for grouping the line pairs on the base image.....	71
Figure 4.5	The constraint of quadrilateral region to reduce the search space .	72
Figure 4.6	(a) Reference line pairs, (b) two quadrilateral regions, (c) the intersection point of $l_1$ and $l_2$ , (d) the epipolar line of intersection and the intersections of candidate pair models that are computed to be less than $T_{epi}$ .....	74
Figure 4.7	The geometrical measures utilized.....	76
Figure 4.8	(a) A reference model, (b) a candidate pair model and its geometrical measures, (c) and (d) normalization with epipolar lines and the normalized measures. ....	77
Figure 4.9	(a) A reference pair and (b) a candidate pair with flanking regions.....	79
Figure 4.10	(a, b) The back projected plane estimated from the line pairs and the intersection point, (c, d) points that are utilized for the one-to-one correspondence based on the height values obtained from the estimated 3D plane.....	80
Figure 4.11	(a, b) Several cases that may violate the plane constraint.....	82
Figure 4.12	(a, b) Corresponding line pairs, (c, d) same line pairs after point-to-point correspondence, (e, f) regions dominated by the line pairs. .	83
Figure 4.13	(a) A reference pair selected from the base image and the intersection point (in red color), (b) the estimated epipolar line of the intersection point on the search image (in dark blue color) and the possible intersection points (in light blue color) of candidate pairs around the epipolar line, and (c) the orthogonal distance between the intersection points and the epipolar line. ....	85

Figure 4.14	The matching ambiguities after pair-wise matching. (a, b) The stereo images and the extracted lines, (c, d) potential ambiguities after the pair-wise matching.....	87
Figure 4.15	Two examples of line segments commonly observed in repetitive patterns and their very similar local neighborhoods. ....	89
Figure 4.16	The original Daisy descriptor.....	90
Figure 4.17	The adaptation of the Daisy descriptor. (a, b) The line segments with point-to-point correspondence and the located Daisy grids to the center of each line segment, (c, d) rotated grids based on the orientation of the line segments, and (e, f) final Daisy grids with adaptive scaling, (g, h) divided Daisy grid points.....	92
Figure 4.18	Line segments extracted from two stereo images. ....	95
Figure 4.19	Left and right stereo images and one of the matching ambiguities due to the perspective viewing and occlusion.....	96
Figure 4.20	A problematic case of the Redundancy measure. ....	97
Figure 5.1	The line segments extracted from two images. The red ones indicate the lines that are aligned ( $\leq 10^\circ$ ) with the epipolar line (flight direction). ....	103
Figure 5.2	The reconstruction of the line segments that is not aligned with the epipolar line.....	104
Figure 5.3	The reconstruction of the line segments that is nearly-aligned with the epipolar line.....	105
Figure 5.4	(a) Confidence regions of the extracted line segments, (b) an example of the generated artificial intersection point and its elliptical confidence region. Entities and confidence regions are shown in red and green colors, respectively. Confidence regions are 15 times exaggerated for visualization.....	106
Figure 5.5	Independent weight curves for each measure. The control parameters ( $\sigma_1$ and $\sigma_2$ ) are (5 and 2), respectively. ....	109
Figure 5.6	Left (a-c) and right (b-d) stereo images. The directional regions utilized during the selection of the best weighted point entities. .	110

Figure 6.1	Vaihingen/Enz test site and the aerial ortho-image (rotated 90° CCW) .....113
Figure 6.2	The Hannover test site and the aerial ortho-images .....114
Figure 6.3	(a) A test building, and two different types of line segments observed. Blue segments illustrate some examples for (b) the step edges, and (c) the intersection edges. The reference LIDAR point cloud is shown in green color.....116
Figure 6.4	The total number of pair-wise relations computed with respect to Proximity measure ( $T_{prox}$ ).....117
Figure 6.5	Six stereo test patches selected from the Vaihingen test site. ....118
Figure 6.6	PDFs of pair-wise constraints and the final similarity function computed for the six Vaihingen test patches.....121
Figure 6.7	(a-c) PDFs of each measure utilized in the precise matching stage, and (d) the number of correct and incorrect matches deleted with respect to different Daisy dissimilarity NNDR thresholds. ....124
Figure 6.8	PDFs of each measure after enforcing the NNDR threshold of 0.1. ....126
Figure 6.9	The results of the performance test based on various hysteresis thresholding combinations for the Vaihingen test site. ....128
Figure 6.10	The results of the proposed approach. The colors green, red, and cyan correspond to correct, incorrect and missed matches, respectively. ....130
Figure 6.11	The final matching performances of the proposed approach computed for the 15 test patches selected from the Vaihingen test site. ....133
Figure 6.12	Six stereo test patches selected from the Hannover test site. ....140
Figure 6.13	PDFs of each pair-wise constraint and the final similarity function computed for the six Hannover test patches. ....141
Figure 6.14	(a-c) PDFs of each measure utilized in the precise matching stage, and (d) the number of correct and incorrect matches deleted with respect to different Daisy dissimilarity NNDR thresholds. ....142

Figure 6.15	PDFs of each measure after enforcing the NNDR threshold of 0.1... .....	143
Figure 6.16	The results of the performance test based on various hysteresis thresholding combinations for the Hannover test site.....	145
Figure 6.17	The results of the proposed approach. The colors green, red, and cyan correspond to correct, incorrect and missed matches, respectively. ....	146
Figure 6.18	The final matching performances of the proposed approach computed for the 15 test patches selected from the Hannover test site. ....	149
Figure 6.19	Performance comparisons of the proposed approach and the Schmid's method for the Vaihingen test patches. ....	153
Figure 6.20	Performance comparisons of the proposed approach and the Schmid's method for the Hannover test patches.....	155
Figure 6.21	Test patches for the evaluation of the proposed reconstruction approach. Left (a-c-e) and right (b-d-f) stereo images. Correct and false matches are shown in green and red colors, respectively....	157
Figure 6.22	3D line segments generated with the method of direct construction (a-d-g) and the proposed reconstruction approach (b-e-h). Block arrows point parts where some of the critical improvements are observed. The LIDAR point cloud overlaid with the lines reconstructed with the proposed approach (c-f-i). Each color in the point cloud belongs to the automatically extracted 3D planes in the vicinity of each line. ....	158
Figure 6.23	The final reconstruction results for the Vaihingen test patches ...	161
Figure 6.24	The final reconstruction results for the Hannover test patches ....	164

## CHAPTER 1

### INTRODUCTION

#### 1.1 Purpose and Scope

Aerial imaging is one of the most common and versatile ways of obtaining information about the objects on the Earth's surface. The information can be quite detailed and of high-quality due to the advanced capabilities of aerial imaging such as the availability of ultra-high-resolution (UHR) cameras (3-10 cm), involved geometric fidelity, broadened spectral sensitivity, permanent recording etc. These special characteristics have turned aerial imagery into one of the most important state-of-the-art data acquisition techniques.

Accurate and reliable knowledge of the terrain surface and/or the height of various off-terrain objects extracted from UHR aerial images are essential for a wide range of applications. Besides, in addition to accuracy and reliability concerns, the reconstructed surface must also be represented in detail to guide further automated image understanding and interpretation tasks. Therefore, the management of height discontinuities turns out to be a very important issue since most of the generated dense surface models suffer from the problem of smoothing caused by the interpolation around the height discontinuities. At this point, feature-based image matching takes an important part, since the height discontinuities in the real world mostly appear as intensity and/or color differences in the images. Managing those differences in different ways determines the type of feature-based matching method to be utilized and finally results in three distinctive types of features that

can be used for matching, i.e., interest points, edges/lines and regions/polygons. The urban areas that consist of various man-made objects are mostly dominated by linear structures. Thus, reliable and accurate reconstruction of 3D edge/line features has a major importance and is a vivid field of research to this day.

Up to now, a significant number of research studies have been completed in the field of line matching from aerial images. However, in a stereo environment, the ambiguity problem of line matching is an issue that remains unsolved. The major problem arises from the lack of measure(s) and/or constraint(s) for line features that are invariant under different viewing conditions. Therefore, although a number of fully automated approaches have been tested so far, achieving acceptable results from those systems are still limited to certain multiple imaging conditions and far below the human perceptual abilities. Besides those matching problems, it is well-known that the reconstruction of straight lines which are nearly parallel to the epipolar line is numerically unstable within a stereo image pair or a single image strip. This is due to the reason that if the angles of lines in image space get closer to the epipolar direction, the projection planes generated from line segments become similar and in the worst case (exact alignment) they turn out to be the same plane. For those cases, the direct construction of 3D lines from the intersection of planes is highly problematic and in some cases the intersection (or the reconstruction) may not be possible.

The work presented in this thesis is a new approach for the matching and 3D reconstruction of straight line features from UHR stereo aerial imagery. The developed scheme aims to provide solutions for the problems stated above and attempts to improve the performances of the matching and reconstruction of the line segments observed in a stereo geometry. The approach consists of three fundamental steps; (i) 2D line extraction, (ii) stereo matching of the extracted line segments, and (iii) stereo reconstruction of the matched line segments (Fig. 1.1). For the implementation of the developed approach, a single stand-alone system that is fully capable to execute the proposed approach is also developed.

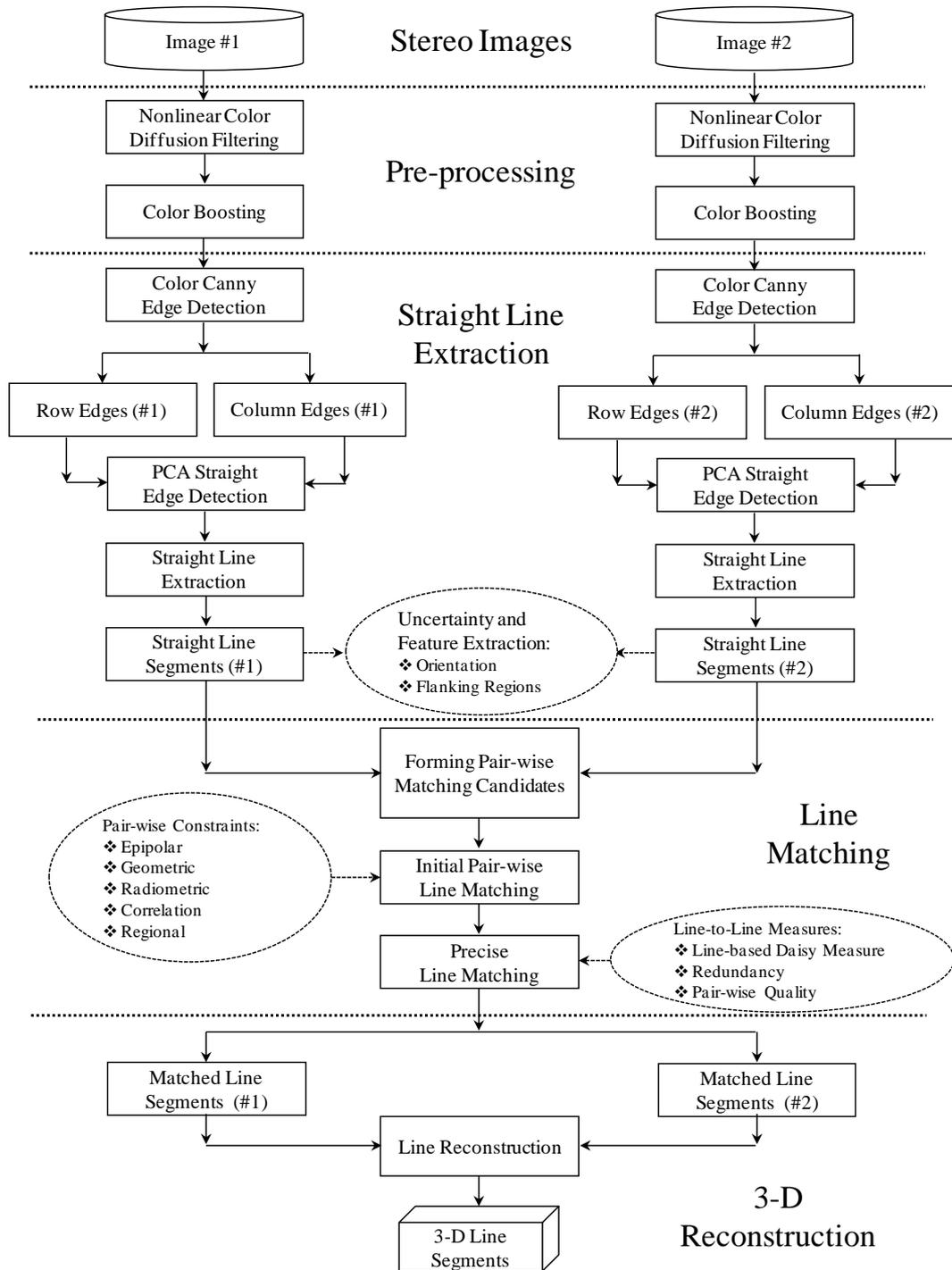


Figure 1.1 Flowchart of the proposed approach.

## 1.2 Contributions

The contributions and innovations of this thesis are:

- Although aerial images have been rich of multispectral information, so far, this fact was almost completely disregarded or not efficiently used during the low level processing such as filtering, edge detection etc. However, multispectral aerial images provide opportunities to extract line features that cannot be detected in the grayscale images due to several reasons, such as low contrast, accidental object alignments etc. Thus, the proposed approach takes full advantage of the existing multispectral information in aerial images in all the steps, especially during pre-processing and edge detection. Thus, even object boundaries that show only a very slight color difference could be detected.
- To accurately describe the straight line segments, a two stage method is proposed, (i) the extraction of straight edge segments with principal component analysis, and (ii) fitting line segments to the extracted straight edge segments. With the improvements of the straight edge detector, the straight line extraction algorithm works quite robustly, even for the areas where an enormous number of edges are found. This offers an opportunity to detect and reconstruct lines that belong to objects and their certain details.
- A new relational approach in which the line correspondences between the stereo aerial images are established in a pair-wise manner is developed. The approach initially generates reference line pairs in the base image and collects all potential matching candidate pairs from the search image. A total of seven unique pair-wise constraints are involved during the matching of line pairs including various new and improved constraints in

different domains such as geometry, radiometry, correlation, regional and epipolarity.

- To assign the final line-to-line correspondences between the stereo images, a precise matching approach is proposed. During the precise matching stage, the matching inconsistencies are eliminated using three novel measures (line-based Daisy, Redundancy, and Pair-wise Quality) and a final similarity voting scheme. At this step, in contrast to previous studies, the line-to-line relations are fully taken into account to solve the matching inconsistencies by imposing the Redundancy measure computed from the entire pair-wise matches.
- A new stereo reconstruction approach is proposed for the line segments that are nearly-aligned ( $\leq 10^\circ$ ) with the epipolar line. The method exploits the redundancy inherent in line pair-relations to generate artificial 3D point entities and utilize those entities during the estimation process to improve the height values of the reconstructed line segments. For the selection of the best point entities for the reconstruction, a new weight function and directional region-based selection method is proposed. Thus, we have a possibility to accurately reconstruct the line segments that are nearly-aligned with the epipolar line.
- A single stand-alone system (within the Matlab 7 environment with graphical user interfaces) that is fully capable of executing the proposed approaches is developed. In addition to the implementation of the proposed approach, the system has additional capabilities such as the processing of the images acquired with both analog and digital frame sensors, performing simultaneous bundle-block-adjustment of frame sensors (with and without self-calibration parameters), etc.

### **1.3 Organization of the Thesis**

This thesis is composed of seven chapters. The next chapter (Chapter 2) provides a review of existing optical airborne cameras and the literature review related to line matching and reconstruction. A short definition and comparison of the vision techniques utilized in this thesis is also presented.

In Chapter 3, the extraction of the 2D line segments is described. The chapter starts with describing the pre-processing techniques utilized. Thereafter, the details of the edge and straight edge detectors are explained. Finally, the extraction of the straight line segments is stated.

The details of the developed line matching approach are provided in Chapter 4. First, the method for the formation of the potential matching candidates is given. Next, the developed matching strategy and the related constraints are clarified. Finally, the proposed precise matching approach is described.

The following chapter (Chapter 5) presents the reconstruction of the matched line segments. First, the generation of the artificial point entities is described. Finally, the details of the approach for the joint height estimation of the point and plane entities are explained.

Chapter 6 involves the information related to the test data utilized, the performance assessment, and the discussion. First, the test sites and the available datasets are described. After that, the matching and reconstruction results of the proposed approach are given. The comparative evaluation of the results of the proposed approach with a state-of-the-art line matching approach is also presented.

In the final chapter (Chapter 7), the conclusions derived from this study and the recommendations that can be useful for further studies are stated.

## **CHAPTER 2**

### **BASICS AND STATE-OF-THE-ART**

#### **2.1 Optical Airborne Cameras**

The standard environment of analogue photogrammetric imaging and processing has been well known and established for decades (Cramer, 2005). However, commercial large-scale digital cameras were introduced to the photogrammetric community just before the beginning of the 21<sup>st</sup> century. Since that time, the digital imaging sensor technologies and the camera systems have significantly evolved. For sure, this technological development has been well issued, followed and reported by the photogrammetric community. An early review of digital frame camera systems can be found in Petrie (2003). Two years after that, Cramer (2005) presented the status of the digital airborne imaging and predicted the role of the digital sensors in the future. Thereafter, the rapid developments in the airborne digital imaging technologies including both digital frame cameras and push-broom line scanners were revealed in another review paper (Petrie and Walker, 2007). Recently, Charles Toth (2009) gave extensive overview lists related to the state-of-the-art airborne optical and LIDAR data collection systems. As a last step of the technological evolution, Jacobsen (2010) pointed out the next generation large-format digital frame cameras that were fully designed with single monolithic charge-coupled device (CCD) sensors. Since UHR aerial images utilized in this thesis are acquired only with large-format frame cameras, in this section, a brief overview and a summary of the current status of the large-scale optical frame airborne cameras is presented.

The analogue cameras are very similar from their design systems using a large format film, and the majority of the analogue photogrammetric images were taken by only two different cameras, namely the RMK, LMK and RC series and their predecessors (Cramer, 2005). In order to achieve high geometric quality, most of the analogue cameras were designed as single-lens concept. In general, 152-mm focal length lenses with a film format size of 230x230 mm are preferred during photogrammetric mapping. Depending on the altitude of the acquisition, 90, 210 and 300 mm focal lengths are utilized (Lillesand, 2005). The angular field-of-view of the lenses is also adaptable and depends on the purpose of the acquisition. Once the aerial images are acquired with analogue cameras, they are converted to raster images by photogrammetric scanners. A common trend for the scanning resolutions is to perform the scanning between 12.5 and 21 microns (Neumann, 2008). The resolution of the scanning and the photo scale both determine the final ground sampling distance (GSD) of each pixel of the output raster aerial images. Despite many years of development, refinement and optimization of performance, the introduction and growing demand of digital cameras has pushed aside analogue cameras. According to Jacobsen (2010), the advantages of the digital cameras are so clear that in several countries, new analogue aerial photos are not accepted anymore. Based on this context, the further production of the analogue cameras has already come to an end.

In the review paper of Petrie and Walker (2007), large-format digital frame cameras are divided into two main categories:

- individual cameras equipped with individual large-format area arrays producing monochrome (black-and-white) frame images
- systems employing multiple medium-format frame cameras which produce sub-images that are combined later using image processing techniques to form a single composite large-format digital frame image

Until recently, all large-format cameras used for civilian purposes belonged to the second category in which several medium-format cameras with relatively small CCD Chip Sizes are utilized to acquire several small images (sub-images) of a region of interest. Those sub-images are then rectified and stitched together to form a single large-format digital monochromatic image (Petrie and Walker, 2007). Color images of the area can be acquired with a second set of medium-format multispectral cameras which are then utilized in the subsequent pan-sharpening process to generate high-resolution color images of the region of interest. Even today, except for the 2<sup>nd</sup> generation of DMC (Digital Mapping Camera) cameras (DMC II), all commercial large-format cameras (Table 1) utilize the same image acquisition logic which can be divided into three distinct groups:

- Closely mounted four tilted cameras synchronically acquire four images with small overlaps to form a single composite panchromatic image (DMC, Trimble AIC x4, DigiCam Quattro)
- In-line mounted four nadir cameras acquire synoptically nine individual images with small overlaps to form a single composite panchromatic image (UltraCam series)
- Side by side mounted two nadir pointing cameras synchronically acquire two images with a small overlap to form a single composite panchromatic image (Dimac Wide+)

The first group cameras utilize four closely coupled cameras together in a block configuration such that all four tilted images are acquired simultaneously and overlap slightly along two of their edges (Petrie, 2009). Figure 2.1a illustrates the position of the four tilted panchromatic heads of the DMC camera. Due to the tilted acquisition with low-oblique imaging, the acquired images must be rectified before the stitching process. The final stitching process that utilizes automatically

Table 2.1 Summary of the large format digital frame cameras (k = 1024 pixels).

<b>Product Name</b>	<b>Focal Length (mm)</b> {Pan + MS}	<b>Pixel Size (μm)</b> {Pan + MS}	<b>Final Image Size (pixel)</b>	<b>CCD Chip Size (pixel)</b> {Pan + MS}	<b>Radiometric Resolution (bits)</b>	<b>Number of Sensors (Pan + MS)</b>	<b>Max Frame Rate (sec / image)</b>
<b>DMC</b>	120 + 25	12 + 12	13,824 x 7,680	{7k x 4k} + {3k x 2k}	12	4 + 4	2.1
<b>DMC II - 140</b>	92 + 45	7.2 + 7.2	12,096 x 11,200	{12.2k x 11.4k} + {6.8k x 6.1k}	14	1 + 4	2
<b>DMC II - 230</b>	92 + 45	5.6 + 7.2	15,552 x 14,144	{15.5k x 14.1k} + {6.8k x 6.1k}	14	1 + 4	1.7
<b>DMC II - 250</b>	112 + 45	5.6 + 7.2	17,216 x 14,656	{17.2k x 14.6k} + {6.8k x 6.1k}	14	1 + 4	1.7
<b>UltraCamD</b>	100 + 33	9 + 9	11,500 x 7,500	{4k x 2.7k}	12	9 + 4	1.3
<b>UltraCamX</b>	100 + 33	7.2 + 7.2	14,430 x 9,420	{5k x 3.4k}	14	9 + 4	1.35
<b>UltraCamXP</b>	100 + 33	6 + 6	17,310 x 11,310	{6k x 4k}	14	9 + 4	2
<b>UltraCamXpWA</b>	70 + 23	6 + 6	17,310 x 11,310	{6k x 4k}	14	9 + 4	2
<b>UltraCamEagle</b>	80 + 27 210 + 70	5.2 + 5.2	20,010 x 13,080	N/A	14	9 + 4	1.8
<b>Dimac Wide+</b>	70/120/210	6 + 6	13,000 x 8,900	{9k x 6.7k}	16	2	2
<b>Trimble AIC x4</b>	60/72/100	6 + 6	17,000 x 12,400	{8.9k x 6.7k}	16	4	2
<b>DigiCam Quattro</b>	80/100/150	6 + 6	18,500 x 12,750	N/A	16	4	1.6

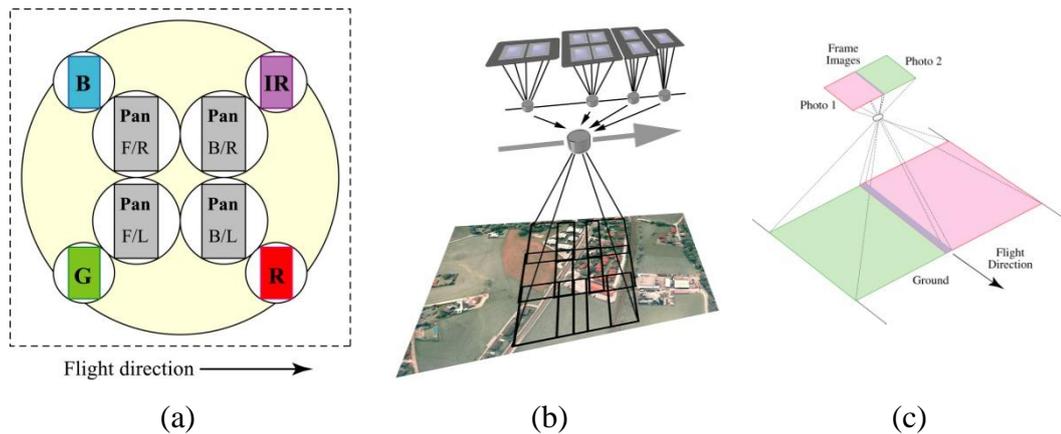


Figure 2.1 Multiple camera image acquisition techniques (Petrie, 2006)

collected tie points converts all four sub-images into a single panchromatic image. In the DMC Camera shown in Figure 2.1a, the four nadir-looking additional small-format cameras positioned in each side are responsible for the acquisition of the reduced-resolution multi-spectral images. However, it should be noted that the other cameras belonging to this group (Trimble AIC x4, DigiCam Quattro) utilize a Bayer pattern for the generation of their true or false color images; thus, they do not involve additional cameras for the multi-spectral image acquisition.

The design concept of the second group cameras (UltraCam series) is composed of a sensor head consisting of eight independent camera cones, four of them contributing to the large format panchromatic image and four contributing to the multi-spectral image. Altogether, the UltraCam sensor head is equipped with 13 CCD sensor arrays (9  $\mu\text{m}$  pixel size for UltraCamD, 7.2  $\mu\text{m}$  pixel size for UltraCamX, 6  $\mu\text{m}$  pixel size for both the UltraCamXp and Xp WA, and 5.2  $\mu\text{m}$  for UltraCamEagle). Four individual camera cones of the panchromatic sensor head have a specific mechanical distance from the camera center which needs to be compensated during image exposure to produce a consistent virtual image (Figure 2.1b). This process is unique and is known as “syntopic exposure”. It exploits the movement of the aircraft in such way that the shutters of the cones are delayed so they don’t open at the same time but at the same position (syntopic) (USGS, 2010). This concept causes all cones to produce sub-images from a single

same physical location in space and thus, a single perspective center is achieved theoretically (Leberl et al., 2003). The final stitching process converts all acquired sub-images into a single image based on a selected master sub-image. Similar to the first group of cameras, automatically generated tie points are required to perform the stitching process which also removes the potential residual errors among different cones.

The camera arrangement of the third group cameras (Dimac Wide+) is based on a twin-design in which two cameras are positioned side-by-side with a nadir-looking geometry (Fig. 2.1c). The CCD area arrays are then offset with respect to the optical axes so that they cover the areas to the left and right of the flight line (Petrie and Walker, 2007). Once again, the full image is formed after a stitching process which utilizes the common overlapping parts between the two images.

All of the three groups described above are very successful in terms of image acquisition with specific characteristics; however, for some cases, they share a common basic problem which occurs due to the process of stitching. The stitching process is based on the tie points collected in an automated manner over the common overlapping parts of the sub-images available (there is a total of 4 sub-images for the first group, 9 sub-images for the second group and 2 sub-images for the last group of cameras). However, the process of stitching requires a textured terrain since a textured environment provides a chance to collect and match large numbers of tie points within the overlap areas. A less-textured terrain leads to a dramatic decrease in terms of the number of tie points collected and therefore, results in lower geometric accuracy for the final stitched image. Furthermore, not surprisingly, in some specific texture-less terrain like a desert area or a water body, it may not be possible to perform the stitching operation. Recently, for the UltraCam images, an interesting solution has been proposed to solve the problem (Ladstädter et al., 2010). The technique is called “monolithic stitching” and the idea is to utilize the geometric information of a single lower resolution color sensor (green) along with the panchromatic sensors in an

integrated adjustment process to collect tie points. However, although the developed technique seems to improve the current stitching performances, Jacobsen (2010) stated that the stitching to a lower resolution reference image is not the optimal solution and contradicts to the syntopic acquisition mode of the UltraCam sensors. It is apparent that the optimal and ultimate solution is not to perform the image acquisition by multiple cameras; thus, all stitching operations can be avoided. But, this certainly requires the manufacturing of single large-format “monolithic CCD” arrays. More recently, this type of large-format CCDs is introduced by the second generation of DMC Cameras (DMC II – 140, 230, and 250). Thus, this group of new commercial sensors utilizes a single panchromatic camera for the image acquisition and resembles the basic optics design principle of the film cameras used for many decades.

In summary, at the moment, current CCD pixel sizes of large-format cameras are on the order of 5  $\mu\text{m}$  and the image size of a single panchromatic frame reached >260 MB. Frame rates of less than 2 seconds give opportunity to maintain high-air speed for high-forward overlap with high resolution. With the availability of single large-format CCD arrays, the geometric quality of the analogue cameras can be achieved by the digital counterparts. To conclude, there is a real competition in terms of the development of the airborne imaging technologies and this is expected to increase in the near-future.

## **2.2 Edge and Line Matching**

The term image matching refers to the automatic generation of correspondences in multiple descriptions of a scene. Once the image-to-image correspondences are established, the 3D information of the scene can be provided. Matching algorithms can be used in different tasks in photogrammetry and remote sensing such as automatic orientation purposes, DSM generation, road and terrain tracking etc. Very detailed overviews and summaries on the topic of image matching can be found in a wide variety of review papers, book chapters and thesis, see e.g.,

Dhond and Aggarwal (1989); Heipke (1996); Heipke (1997); Schenk (1999); Manual of Photogrammetry (2004); Zhang (2005). Since this thesis is devoted to the matching of line features, this section gives a review about the previous studies on the subject of line matching from aerial and/or satellite images (Table 2.2). The first part examines the previous studies in an article by article manner, which investigates and summarizes each article based on their certain key aspects such as the methodology proposed, the data used, and the results reported. The second part is devoted to the discussion of the previous studies in a categorizing manner, which classifies previous studies based on several aspects such as the trends followed, data used, etc.

### **2.2.1 Related Work**

In an early work, Ohta and Kanade (1985) presented a stereo matching algorithm that utilized the edges as fundamental elements to be matched, and employed two different search techniques for the matching procedure. The first technique was the inter-scan line search for possible correspondences of connected edges in the right and left images, and the second technique was the intra-scan line search for correspondences of edges on each scan line pair. In their method, dynamic programming was utilized for the implementation of the search techniques which proceeded simultaneously in two ways: the former supplies the consistency constraints to the latter while the latter supplies the matching score to the former. A final interval-based similarity metric was used to compute a matching score. The method was tested on two urban aerial images, and found to be satisfactory in terms of matching the edges in a stereo environment.

Herman and Kanade (1986) proposed an incremental system for the reconstruction of 3D scenes. In their work, the stereo matching was performed by the junction features extracted from the intersection of the line segments, and sparse 3D wire-frames were generated. Thereafter, linear structures were extracted from monocular image analysis, and based on several task-specific assumptions

Table 2.2 Summary of the previous work on the topic of line matching using aerial and/or satellite images. Terms “V” and “O” are acronyms for “Vertical” and “Oblique” images, respectively.

<b>Previous Work</b>	<b>GSD (cm)</b>	<b>Num. of images</b>	<b>Data Source</b>	<b>Utilized Radiometry</b>	<b>Supplementary Data/Method</b>
Ohta & Kanade (1985)	N/A	2	Analog (V)	Grayscale	-
Herman & Kanade (1986)	N/A	2-3	Analog (V)	Grayscale	Junctions by Image Matching
Greenfeld & Schenk (1989)	N/A	2	Analog (V)	Grayscale	-
Mohan et al. (1989)	N/A	2	Analog (V)	Grayscale	-
Hoff & Ahuja (1989)	N/A	2	Analog (V)	Grayscale	-
Cochran & Medioni (1992)	N/A	2	Analog (V)	Grayscale	-
Roux & McKeown (1994)	50	4	Analog (V & O)	Grayscale	Corners by Image Matching
Jordan & Cocquerez (1995)	85	2	Analog (V)	Grayscale	-
Bignone et al. (1996)	7.5	4	Analog (V)	Color	-
Henricsson & Baltsavias (1997)	7.5	4	Analog (V)	Color	-
Schmid & Zisserman (1997)	8.5	2-3	Analog (V)	Grayscale	-
Atalay & Yilmaz (1998)	N/A	2	Analog (V)	Grayscale	-
Collins et al. (1998)	31-110	>2	Analog (V & O)	Grayscale	-
Henricsson (1998)	7.5	4	Analog (V)	Color	-
Moons et al. (1998)	8	3-6	Analog (V)	Color	-
Baillard et al. (1999)	8.5	>3	Analog (V)	Grayscale	-
Baillard & Dissard (2000)	16-103	2	Analog (V)	Grayscale	Area-based matching
Park et al. (2000)	N/A	2	Analog (V)	Grayscale	-

Table 2.2 (continued)

<b>Previous Work</b>	<b>GSD (cm)</b>	<b>Num. of images</b>	<b>Data Source</b>	<b>Utilized Radiometry</b>	<b>Supplementary Data/Method</b>
Schmid & Zisserman (2000)	8.5	2-3	Analog (V)	Grayscale	-
Scholze (2000)	8	4	Analog (V)	Color	-
Scholze et al. (2000)	8	4	Analog (V)	Color	-
Shao et al. (2000)	N/A	3	Analog (V)	Grayscale	-
Zhang & Baltsavias (2000)	20-40	2	Analog (V)	Color	-
Cheng et al. (2001)	31-110	>2	Analog (V & O)	Grayscale	-
Heuel & Förstner (2001)	N/A	>2	Analog (V)	Grayscale	Corners by Image Matching
Leloglu (2001)	7.5-8.5	3-4	Analog (V)	Color	Planar Surfaces
Noronha & Nevatia (2001)	15-130	2-4	Analog (V & O)	Grayscale	Junctions by Image Matching
Chehata et al. (2002)	50-60	2	Analog (V)	Grayscale	DSM
Jung et al. (2002)	20	9	Digital (V)	Grayscale	DSM
Taillander & Deriche (2002)	20	6	Digital (V)	Grayscale	-
Elaksher et al. (2003)	12	4	Analog (V)	Grayscale	Regions by Image Matching
Jung & Paparoditis (2003)	20	9	Digital (V)	Grayscale	DSM
Beder (2004)	14.3	7	Analog (V)	Grayscale	-
Kim & Nevatia (2004)	25	4-7	Analog (V)	Grayscale	Junctions by Image Matching
Zhang (2005)	5.7	3-6	Digital (V)	Grayscale (Red Band)	Point Cloud by Image Matching
Zhongliang & Zhiqun (2008)	N/A	2	Analog (V)	Grayscale	-
Woo et al. (2009)	7.5-100	2	Analog Digital (V)	Grayscale	-

Table 2.2 (continued)

Previous Work	GSD (cm)	Num. of images	Data Source	Utilized Radiometry	Supplementary Data/Method
Habib et al. (2010)	6	2	Analog (V)	Grayscale	Point Cloud by LIDAR
Xiao et al. (2010)	10-16	8	Digital (O)	Color	-

the final 3D wire frames were generated. The intermediate and final visual results are illustrated for a set of analogue aerial images.

A different study that implemented stereo matching by means of edge-based features was performed by Greenfeld and Schenk (1989). They presented two methods for describing the edge features; (i) a method based on a polygonal approximation (vertices and legs) and (ii) a method derived from  $\Psi$ - $S$  curve which is a functional representation of a line where the arc length  $S$  is the parameter of the tangent  $\Psi$ . During the matching of the edges, first, an initial matching list was established by taking into account four similarity measures (angle and orientation of the polygon vertices, steepness and sign of the zero-crossings). Next, the ambiguities of the initial matching list were resolved by a consistency check that involved a voting based on vertical disparity, azimuth and distance between pairs of matched vertices. The result of the matching was assessed visually and reported as successful.

Mohan et al. (1989) described a method which utilized linear segments to improve the performance of the stereo matching. The stereo matches were collected from the algorithm developed by Ohta and Kanade (1985). After that, they presented a post-processing method which rejected the disparities violating the constraint of figural continuity. They computed a linear function describing the disparity change along the linear segments, and used the function to correct the wrong

disparities due to incorrect matches. Furthermore, a total of four cost functions were also proposed to quantitatively analyze the quality of the matching.

An integrated approach for the reconstruction of surfaces from stereo matching was performed by Hoff and Ahuja (1989). The aim of the approach was to enforce a surface smoothness constraint during the matching stage which was based on a multi-resolution strategy. The boundary information is then extracted by locating the ridges and contours. The evaluations of the approach were performed for different sets of images including a single stereo aerial image. As a result, the problems and the advantages of the proposed approach were stated.

Cochran and Medioni (1992) proposed an approach to combine the information extracted from area-based and edge-based approaches. The approach utilized manually registered three-level image pyramids to provide dense depth maps of the regions of interest. The normalized cross-correlation and a set of constraints (left-right checking, ordering, and no-isolated pixels) were imposed to improve the performance of the area-based method. A local interpolation was then performed and edge-based approach was used to extract the discontinuities to remove the smoothing effects from the final disparity map. Only visual results were given for a set of six close range images and one aerial image.

Roux and McKeown (1994) proposed a feature based matching approach to be used for the extraction of buildings from multiple aerial images. In their approach, the corner features were used as a fundamental cue for matching, and the edge information was integrated to infer 3D segments. For the matching process, they fully utilized the knowledge about the imaging geometry and acquisition parameters to provide several geometric constraints (epipolar, height and orientation). Their study was also one of the first (studies) that incorporated multiple aerial images having large amounts of obliquity. The results of the matching process were evaluated based on a visual assessment.

Jordan and Cocquerez (1995) presented a matching approach to guide the 3D description of urban areas from stereo aerial images. The matching process was initialized with a search interval defined along the epipolar line based on a function of the altitude variation in a scene. Next, for the edge segments, a search depending on the similarity of the alpha parameter of the Canny-Deriche (Deriche, 1987) edge detector was performed. In addition, the ordering constraint was imposed to further improve the performance of the matching process. Once the search was completed, the consistency of matching was further checked to eliminate isolated and aberrant matches. The approach was tested in two urban test sites, and the matching results were illustrated. Although the numerical results for the reconstruction stage were provided, no explicit matching performances were given.

One of the early works that investigated the color information during the matching process was performed by Bignone et al. (1996). Later, the same matching strategy was also followed in several different studies (Henricsson and Baltsavias, 1997; Henricsson, 1998). The idea was to automatically match and reconstruct 3D segments to guide the automated reconstruction of buildings. First, the edge information was extracted from a specific image and the related matches were collected from other available images based on an edginess measure (based on image gradients) computed along the epipolar lines. The approach used the advantages of multiple (a total of four images) image geometry as well as photometric edge attributes (luminance and chromaticity) to perform high-level segment matching. The performance of the approach was evaluated on a residential test site and the results contained a number of incorrect matches. Most of those incorrect matches were later removed after performing the coplanar grouping of the 3D segments for the building reconstruction task.

A rigorous method for the matching of line segments acquired in stereo and/or triple imaging combinations was developed by Schmid and Zisserman (1997). In their method, the epipolar geometry was successfully utilized to provide point-to-

point correspondences along the line segments. To exploit the intensity neighbourhood of the segments, direct and warped correlation measures were proposed around the line neighbourhoods observed by short and long range motions, respectively. For the matching of line segments in three-views, they proposed two different methods both of which fully utilize the trifocal tensor. In their paper, explicit matching results were given for image pairs and triplets, and very high rates ( $> 95\%$ ) of matching accuracies were computed. Later, their algorithm was extended to multiple views ( $> 3$ ) by Baillard et al. (1999). First, they collected the line matches in a single strip triple combination and next, the matches were verified in other images available in the adjacent strip(s). Their multiple view approach also revealed very good performances ( $> 98\%$ ) for the matching lines in three or more views. The same approach for the line segments and its extension to curve matching can also be found in detail in the work performed by Schmid and Zisserman (2000).

Atalay and Yilmaz (1998) proposed a two step feature based matching algorithm based on linear features. In the first step, the rotation angle ( $\Delta\kappa$ ) between the stereo images was computed and the lines extracted from one of the images were rotated based on the other image with the computed rotation angle. In the second step, the line geometric attributes were used and the matching was performed by relaxation. The conditions on which the matching algorithm was intended for was also clearly provided in the paper. The algorithm was tested with a single aerial image pair and the final matching performance was computed to be nearly 87%. It was also pointed out by the researches that most of the mismatches occurred due to incorrectly matched short line segments.

Collins et al. (1998) designed and developed a system to reconstruct buildings from multiple overlapping aerial images. First an automated building detector was run on one of the images to hypothesize potential building rooftops and next, those rooftop polygons were matched in other overlapping images with a multi-baseline stereo matching algorithm. The basic primitives of matching were the

line segments that formed the polygons. The matching ambiguities were resolved by multi-image verification supported by epipolar geometry. The system was evaluated using both nadir and oblique images with varying numbers of images, and only the reconstruction performances were stated. According to the results, 3D building corner positions were recovered to well within a meter of accuracy, with height being estimated more accurately than horizontal position.

Moons et al. (1998) presented an approach for automated modeling and 3D reconstruction of urban buildings from high-resolution multiple aerial images. During the roof modeling process, straight line segments were selected as image features. Epipolar geometry was utilized to collect line matches and the trifocal constraints were applied to eliminate the matching ambiguities. It is noteworthy to state that only the matches visible in at least three images were accepted and utilized for the subsequent polygon grouping stage. No explicit performances of line matching were stated in the paper.

In the study of Baillard and Dissard (2000), area-based and edge-based matching approaches were both utilized to generate a dense DSM. The approach was initialized with an edge matching step. The matching was performed by dynamic programming based on a cost function. Three attributes were utilized during edge-matching; orientation of the image gradient and two intensity values on the sides of the edges. In the next step, the edges were used as supplementary information for the area-based matching procedure to provide a better dense matching result. No explicit numbers of the correct and incorrect edge matches were stated in the paper.

Park et al. (2000) proposed an eigenvector-based line feature matching approach that was based on the geometric relations between pairs of line segments. First, a preliminary correspondence test consisting of a total of four geometric measures was forced to restrict and reduce the number of matching possibilities. Second, a modal analysis with two non-directional measures was utilized to yield a final

dissimilarity measure and to determine the quantitative degree of the dissimilarities between the two pairs of line segments. The method was tested on several synthetic, close-range and aerial stereo images, and as a result, the authors claim that the performance of matching bears a resemblance to human perception.

Shao et al. (2000) described an approach for edge-based matching of line segments from multi-view aerial imagery. The approach consisted of three steps. In the first step, initial edge matches were collected by means of epipolar and trifocal constraints. Next, the number of matches was extended through segment prediction which utilized a local affine model with a certain error space. Finally, a consistency checking by means of a relaxation process was applied. The assessments on an aerial image indicated that the proposed approach yielded reliable matches and a successful recovery.

A multi-image matching algorithm that significantly benefits from the multispectral nature of the aerial images was proposed by Scholze (2000) and Scholze et al. (2000). In the algorithm, chromatic similarity measures were combined with cross-correlation measure (Schmid and Zisserman, 1997) to reduce the matching ambiguities. The approach also utilized multi-view geometric constraints to further restrict and eliminate the mismatches. The reconstruction of the line matches collected from multiple images was performed by the singular value decomposition (SVD) technique. The tests of the approach were carried out on an aerial image acquired with a four-way overlap and it was stated that the presented method produced a highly reliable set of 3D line segments.

In a different work, Zhang and Baltsavias (2000) proposed a line matching approach for the reconstruction of road networks from aerial images. Their method utilized similarity measures including photometric edge attributes along with the geometric information between the line segments. The epipolar constraint was applied to reduce the search space. The final matching list was achieved through structural matching with probability relaxation. The experiments were

performed on a number of image patches extracted from stereo aerial images and the matching approach was found to be very reliable with a high success matching rate.

Cheng et al. (2001) presented a study that involved the topic of matching and reconstruction of point and line features. Two different affinity measures were proposed in the study, and those measures were also utilized to determine the final matches in a weighted bipartite graph matching. The matching for the line features required at least three images since the affinities for the line segments could not be computed in a stereo environment. Furthermore, during the line matching, a constraint that dealt with the common parts of the line segments between the multiple images was also imposed. The performance tests were performed on a number of synthetic and real images; however, for the aerial dataset, only visual matching results were provided.

Heuel and Förstner (2001) presented a method for the matching of line segments in multiple views. In their work, the fundamentals for the optimal reconstruction of the 3D line segments from the matched features were also described. Their matching method was based on multi-view geometrical constraints with error propagation and hypotheses testing. Corner features were also integrated into the matching stage and a final grouping phase was performed along with the line and corner features. The tests of the method were conducted for a test site that was visible in four different aerial views and results of the method were found to be reasonable for both the matching and grouping stages.

A 3D reconstruction approach that deals with dense and complex environments was proposed by Leloglu (2001). In his work, for the matching and reconstruction, area and feature based matching approaches were simultaneously taken into account. In the feature matching part, the similarities were computed by a similarity function that also employs planar surfaces. The assessment of the similarities was performed by evaluating three geometrical attributes (orientation,

length and distances). The line matching was accomplished after projecting the line segments onto the planes defined by the planar surfaces. The approach was tested with two different test datasets and explicit numerical and visual results for the reconstructed surfaces were provided.

Noronha and Nevatia (2001) proposed a hierarchical hypothesize and verify strategy for the detection and reconstruction of buildings. The general strategy used was, first, to generate parallels from the line segments, then, U-shapes from the parallels, and finally, rectangular features from the U-shapes. The generated features at each step were matched in the hierarchy using the available multi-images. The line matches were collected with a single epipolar constraint and the junction features were utilized for disambiguation along with a set of geometric constraints (orthogonality, trinocular etc.). The aerial images were acquired from both nadir and oblique angles. The paper presents explicit results for the final building detection and reconstruction, but no specific matching performances were provided. Later, their line matching algorithm was also implemented by Kim and Nevatia (2004).

In the work performed by Chehata et al. (2002), 3D line segments were utilized to recognize building objects. During the line matching stage, epipolar information was combined with an external DSM to reduce the search space and to limit the matching complexity. A geometric constraint and the photometric neighborhoods of the line segments were evaluated for eliminating the wrong matches. A final matching score was computed from the geometric and photometric information, and the matching was performed based on a winner-takes-all scheme. The numerical results of the approach for a single test site were provided and the results revealed a number of mismatches.

Jung et al. (2002) and Jung and Paparoditis (2003) proposed an edge matching technique to be used for the matching and reconstruction of both straight and curved linears. The method used the epipolar constraint to collect potential

matching edges from the search images. Next, robust bundle estimation was applied to test and analyze the distances of each image ray to the reconstructed candidate 3D object points. The image rays obtaining large distances were eliminated from the matching list, and only the matches observed in more than three images were accepted as a potential matching edge. Thereafter, the 3D tangent directions of each edge were computed and analyzed to further eliminate the incorrect matches. Finally, by means of an external robust DSM, the remaining mismatches were eliminated. A total of nine overlapping aerial images were utilized during the matching process, and the final matches were found to be successful.

Taillander and Deriche (2002) presented a method for the reconstruction of 3D segments from multiple images. They performed the matching of line segments in object space through a plane sweeping method. Thereafter, the matched line segments were reconstructed and the residuals of the reconstruction process were utilized as a geometric criterion to evaluate the correctness of the correspondences. In addition, a unicity criterion was applied to further prune the set of possible correspondences. Both synthetic and real images were used to assess the approach. For the tests that involved real aerial datasets, a total of six overlapping images were utilized and promising matching results were achieved.

A roof boundary extraction technique from multiple aerial images was proposed by Elaksher et al. (2003). The idea was to first extract and divide regions into two classes (roof and non-roof). Next, the region correspondences were established over the images with the aid of regional constraints such as region size, region shape, and region intensity values and an epipolar constraint. Thereafter, the line segments within the corresponding regions were matched over available images. Four-way overlapping images were used to evaluate the algorithm and explicit reconstruction performances were provided. It was also stated that the algorithm provided successful matching results for the line matching.

A unified framework for the automatic reconstruction of point and line features from multiple images was presented by Beder (2004). Statistical geometric properties of the extracted features were utilized in image domain and the framework utilized graphs deduced from relational geometric properties. The features that were visible by a minimum number of images across all views were finally accepted as correct matches. A total of seven overlapping images were used to test the framework, and based on the results, the author claimed that the geometric information from the images was sufficient to establish successful matches over multiple images without requiring any assistance of the radiometric information.

In a different work, Zhang (2005) developed an edge matching approach for UHR linear array images for the automated generation of DSMs. In his work, epipolar geometry, average intensities around line neighborhoods (edge signs), and line orientations in image space were used to restrict the matching. In addition, a-priori matched point features were also integrated into the edge matching stage to further reduce the matching ambiguities. Final matches were assigned after consistency checking with iterative probabilistic relaxation. Successful matching and reconstruction results were provided for a set of images acquired with linear arrays, including images from both satellite and airborne platforms.

Zhongliang and Zhiqun (2008) implemented a method for matching line segments in stereo geometry. They adopted a hierarchical straight line matching strategy. First, an initial matching was performed to restrict and reduce the matching possibilities. Next, the matching was completed by weighting a number of geometric and radiometric measures including various flanking region measures. The length, direction, position, intensity, sharp degree, brightness, darkness, contrast measures are involved during the matching stage and computed for each line and the flanking regions. The results were presented for three test sites, and an average of 89% correct matching rate was computed.

Woo et al. (2009) proposed a method for line matching based on geometric and intensity information extracted for each line segment. A set of line attributes (intensity, length, orientation, end-points) and constraints (epipolar and figural continuity) were implemented during the matching phase. Thereafter, those attributes were utilized with a multi-thresholding scheme in which the matching thresholds for each attribute were iteratively modified. Stereo aerial and satellite images were used to test the method and a final matching correctness ratio of 96% was reported.

A recently proposed approach that integrated LIDAR derived information to the line matching process was performed by Habib et al. (2010). In the approach, the line segments were matched after projecting them onto the LIDAR planes derived from a segmentation procedure. The matching was performed by taking into account the projected line segments with three geometric constraints (angular deviation, normal distance, and overlap). Finally, those matched segments were utilized to acquire precise boundaries of the building objects. Although specific results for the building extraction and reconstruction were provided in the paper, the matching performances were not explicitly stated.

In a different recent work, Xiao et al. (2010) proposed a methodology for the automatic detection of buildings from multiple oblique aerial images. The first step of the methodology was to robustly acquire 3D line segments from the overlapping eight oblique views. To do that, line hypotheses were generated by means of stereo images, and after that, the verification of the stereo hypotheses were performed with the aid of other images in a geometric manner. A matching line segment was accepted if that segment was observed in at least six images. Although detailed numerical results for the building detection task were provided in the paper, the line matching performances were only evaluated qualitatively.

### 2.2.2 Discussion of the Related Work

Up to now, a significant number of research papers have been published in the field of line matching from aerial images. A useful classification of existing line matching approaches was proposed by Schmid and Zisserman (1997). They divided the line matching algorithms into two types, (i) those that match individual line segments, and (ii) those that match groups of line segments. In any case, the search space for matches has to be pruned in some way in order to limit the matching complexity. For most of the studies, basic geometric parameters of line segments such as orientation, length, mid-point, etc. are involved to filter the set of correspondence hypotheses; however, probably the most preferred constraint is the quadrilateral constraint generated using the epipolar geometry (Roux and McKeown, 1994; Collins et al., 1998; Moons et al., 1998; Heuel and Förstner, 2001; Noronha and Nevatia, 2001; Chehata et al., 2002; Kim and Nevatia, 2004; Suveg and Vosselman, 2004; Woo et al., 2009). Some studies also investigated the radiometric information around the line segments (Bignone et al., 1996; Schmid and Zisserman, 1997; Henricsson, 1998; Baillard et al., 1999; Scholze et al., 2000; Schmid and Zisserman, 2000; Shao et al., 2000; Zhang and Baltsavias, 2000; Chehata et al., 2002; Zhongliang and Zhiqun, 2008; Woo et al., 2009) or the information extracted from image gradients (Bignone et al., 1996; Baillard and Dissard, 2000). Additional constraints such as surface smoothness (Hoff and Ahuja, 1989), uniqueness (Suveg and Vosselman, 2004), ordering (Jordan and Cocquerez, 1995; Suveg and Vosselman, 2004) and figural continuity (Mohan et al., 1989; Zhang, 2005; Woo et al., 2009) can also be included; however, for a stereo line matching problem, these constraints are not sufficient to solve the image to image multi-correspondence problem. Thus, additional effort has been spent on different algorithms to select the best line correspondences. For example, dynamic programming (Ohta and Kanade, 1985; Yip and Ho, 1996; Baillard and Dissard, 2000), weighted criterion functions (Henricsson, 1998; Zhongliang and Zhiqun, 2008), modal analyses (Park et al., 2000), graph based approaches (Cheng et al., 2001; Beder, 2004) and probabilistic relaxation (Atalay

and Yilmaz, 1998; Shao et al., 2000; Zhang and Baltsavias, 2000; Zhang, 2005) are among those approaches.

So far, in a stereo environment, the ambiguity problem of line matching is an issue that remains unsolved. The major problem arises from the lack of measure(s) and/or constraint(s) for line features that are invariant under different viewing conditions. Furthermore, the stereo matching and reconstruction of the line segments that are nearly-aligned with the epipolar line is also very problematic since a point-to-point correspondence is not reliable for those cases. Therefore, the general attempt is to strengthen the geometrical constraint by integrating one or more additional views (Roux and McKeown, 1994; Bignone et al., 1996; Henricsson and Baltsavias, 1997; Schmid and Zisserman, 1997; Collins et al., 1998; Henricsson, 1998; Moons et al., 1998; Baillard et al., 1999; Scholze, 2000; Heuel and Förstner, 2001; Noronha and Nevatia, 2001; Jung et al., 2002; Elaksher et al., 2003; Jung and Paparoditis, 2003; Kim and Nevatia, 2004; Taillandier and Deriche, 2004; Zhang, 2005; Xiao et al., 2010). Several others utilized external DSMs (Chehata et al., 2002; Jung and Paparoditis, 2003; Taillandier and Deriche, 2004), point/corner/junction features (Herman and Kanade, 1986; Roux and McKeown, 1994; Heuel and Förstner, 2001; Noronha and Nevatia, 2001; Kim and Nevatia, 2004; Zhang, 2005), surfaces/regions (Stilla and Michaelsen, 1997; Leloglu, 2001; Elaksher et al., 2003) or LIDAR data (Habib et al., 2010) for both reducing the search space and filtering out the matching ambiguities. Nevertheless, the final matching performance of those algorithms is highly dependent and determined by the efficiency and the quality of the auxiliary information. On the other hand, the probabilistic relaxation based methods (Zhang and Baltsavias, 2000; Zhang, 2005) utilize the predefined local neighborhood information which mostly suffer from the piecewise smoothness constraints involved. Inevitably, smoothing based on the local neighborhood violates the standpoint of height discontinuity (except artificial edges such as shadows etc.) of the edges and the subsequent line matching.

To summarize, most of the previous work related with stereo line matching relies on various descriptors specialized for one-to-one line matching. The aim was always towards to reduce the ambiguity problem by integrating additional view(s) to the matching stage. However, the topic of automated line matching has proved to be a challenging task even for images formed under perfect imaging conditions. Although a number of fully automated approaches have been tested so far, achieving acceptable results from those systems are still limited to certain multiple imaging conditions and far below the human perceptual abilities.

### **2.3 A Short Comparison of the Vision Techniques Utilized**

In this thesis, during the matching step (estimation of epipolar lines, stereo intersection etc.), the well known non-linear (or linearized) photogrammetric concepts are applied. On the other hand, during the reconstruction step (2D and 3D point generation, 3D line generation, error propagation etc.), concepts of computer vision are applied. Both approaches are used and implemented by different disciplines and therefore, have different benefits and drawbacks. Fundamental photogrammetric concepts have been known for a long time and are well-described in Manual of Photogrammetry (1980), Kraus (1993), Wolf and Dewitt (2000), Mikhail et al. (2001), and Manual of Photogrammetry (2004). Alternatively, the computer vision concepts are relatively newly implemented; essentials of the computer vision based techniques can for example be found in Faugeras (1993), Hartley and Zisserman (2001), Heuel (2004), Manual of Photogrammetry (2004), and Paragios et al. (2005). The comparison between the two concepts can also be found in (Mundy, 1993; Förstner, 2002; Ressel, 2004; Manual of Photogrammetry, 2004; Förstner, 2009). An interested reader may refer to those references; here, we only very briefly review and compare those vision techniques.

The primary objectives of photogrammetry are the extraction of topography related information and the identification of various objects from terrestrial

landscapes with satisfying final accuracy in all three dimensions. Depending on the type of sensor (frame, linear array, etc.), the geometric camera modeling may vary; however, today, most of the aerial cameras utilize a perspective camera model in which the light rays intersect at a single point called projection center. Due to the central projection, the fundamental collinearity equations can be used to relate the image space and the object space (Manual of Photogrammetry, 2004):

$$\begin{aligned} x - x_0 &= -f \frac{R_1}{R_3} = F_1 \\ y - y_0 &= -f \frac{R_2}{R_3} = F_2 \end{aligned} \tag{2.1}$$

$$\begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} = R \times \begin{bmatrix} X - X_L \\ Y - Y_L \\ Z - Z_L \end{bmatrix} \tag{2.2}$$

where;  $x$  and  $y$  are the image space coordinates of a point,  $X, Y, Z$  are the object space coordinates of the point,  $X_L, Y_L, Z_L$  are the object space coordinates of the perspective centre,  $f$  is the focal length of the sensor,  $x_0$  and  $y_0$  are the coordinates of the principal point, and  $R$  is the rotation matrix that represents the  $\omega, \varphi, \kappa$  rotations of the image coordinates with respect to ground coordinates.

Eq. 2.1 assumes perfect conditions for relating an image space with the corresponding object space. However, during image acquisition, systematic errors may arise from a number of sources (Manual of Photogrammetry, 2004). To remove systematic errors, additional parameters (APs) can be introduced to collinearity equations (Fraser, 1997; Manual of Photogrammetry, 2004; Habib et al., 2002). After including the APs in the collinearity model, Eq. 2.1 becomes

$$\begin{aligned}
x - x_0 + \Delta x &= -f \frac{R_1}{R_3} \\
y - y_0 + \Delta y &= -f \frac{R_2}{R_3}
\end{aligned}
\tag{2.3}$$

where;  $\Delta x$  and  $\Delta y$  represent the correction functions for image coordinates. Of the AP model categories, the physical models seek the most appropriate parameters for computing the correction functions of the image coordinates based on four principal sources of distortions. These distortions include the symmetric radial distortion, decentering distortion, image plane unflatness, and in-plane image distortion (Fraser, 1997). At any point in the image, the net image displacement (Eq. 2.4) will amount to the cumulative contribution of these distortions (Manual of Photogrammetry, 2004).

$$\begin{aligned}
\Delta x &= \Delta x_r + \Delta x_d + \Delta x_u + \Delta x_f \\
\Delta y &= \Delta y_r + \Delta y_d + \Delta y_u + \Delta y_f
\end{aligned}
\tag{2.4}$$

In Eq. 2.4, the subscripts  $r$ ,  $d$ ,  $u$ , and  $f$  represent, the radial distortion, decentering distortion, image plane unflatness, and the in-plane distortion, respectively. The sources and the formulation of these distortions are well explained in Fraser (1997), Manual of Photogrammetry (2004), and Poli (2005). As an example, the complete AP formation including the corrections for the interior orientation parameters can be given as:

$$\begin{aligned}
\Delta x &= -x_0 - \frac{\bar{x}}{f} \Delta f + \bar{x} r^2 K_1 + \bar{x} r^4 K + (2\bar{x}^2 + r^2) P_1 + 2P_2 \bar{x} \bar{y} - A_1 \bar{x} + A_2 \bar{y}_2 \\
\Delta y &= -y_0 - \frac{\bar{y}}{f} \Delta f + \bar{y} r^2 K_1 + \bar{y} r^4 K_2 + (2\bar{y}^2 + r^2) P_2 + 2P_1 \bar{x} \bar{y} + A_1 \bar{y}
\end{aligned}
\tag{2.5}$$

$$r^2 = \bar{x}^2 + \bar{y}^2 = (x - x_0)^2 + (y - y_0)^2$$

where;  $\Delta f$  is the correction to focal length,  $K_1, K_2$  are the symmetric radial lens distortions, and  $P_1, P_2$  are the decentering distortions parameters. The terms  $A_1$  and  $A_2$  introduce the in-plane distortions and represent the coefficients of the affine distortion model. In Eq. 2.5,  $A_1$  and  $A_2$  were designed in a way that they eliminate the correlations between the other APs. This is done by eliminating the shifts of the affine model since the coefficients are perfectly correlated with the coordinates of the principal point (Habib et al., 2002).

As can be seen easily in Eq. 2.1, regardless of any AP, the collinearity equations are inherently nonlinear. Therefore, from a photogrammetric point of view, since the equations of the central projection are non-linear, a linearized form of those equations is required in parameter estimation, and any adjustment requires iterative solutions with appropriate approximate initial values.

Alternatively, computer vision approaches also obeys the central projection principle. Projective geometry is very useful during the representation of the computer vision problems. In projective geometry, an object point in 3D space with homogeneous coordinates  $\mathbf{X} = (x, y, z, t)^T$  represents a 3D point with Euclidean coordinates  $\mathbf{X} = (x/t, y/t, z/t)^T$ , and an image point in 2D space with homogeneous coordinates  $\mathbf{x} = (u, v, w)^T$  represents a 2D point with Euclidean coordinates  $\mathbf{x} = (u/w, v/w)^T$ . From this definition, the photogrammetric collinearity equations given in Eq. 2.1 can be replaced by a linear representation for the central projection, and the projection from the object point  $\mathbf{X}$  to the image point  $\mathbf{x}$ , both represented by homogenous vectors, can be obtained by a matrix multiplication with a single point projection matrix ( $\mathbf{P}$ ) (Manual of Photogrammetry, 2004):

$$\mathbf{x} \cong \mathbf{P} \mathbf{X} \quad (2.6)$$

The 3x4 point projection matrix can be related to the interior and exterior orientation parameters of the camera used in Eq. 2.1 and Eq. 2.2 (Manual of Photogrammetry, 2004):

$$\mathbf{P} = \mathbf{KR}[I - \mathbf{X}_L] \quad (2.7)$$

where,  $\mathbf{X}_L$  is object space coordinates of the perspective centre, and  $\mathbf{R}$  is the rotation matrix, and  $\mathbf{K}$  represents the combined calibration matrix including the three interior orientation parameters ( $f, x_0, y_0$ ) and as well as the scale difference  $m$  and the shear  $s$ . With this definition, the explicit mapping with the elements of  $p_{ij}$  of matrix  $\mathbf{P}$  turns out to be:

$$\begin{aligned} x &= \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}} \\ y &= \frac{p_{21}X + p_{22}Y + p_{23}Z + p_{24}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}} \end{aligned} \quad (2.8)$$

Apparently, if the APs are introduced in the collinearity model as in Eq. 2.3, the projection model cannot be handled in a linear way, since the distortions have non-linear characteristics. Furthermore, as can be easily seen in Eq. 2.5, the distortions of APs depend on the specific image positions within the entire image. According to Förstner (2004), there are two alternative solutions to include APs into the projective model. One solution is to assume locally constant distortions for specific window sizes which results in a local projective model. The second solution is to perform the projection in a two step computation:

$$\mathbf{x} \cong \mathbf{P} \mathbf{X} \quad \rightarrow \quad \mathbf{x}^g \cong \mathbf{K}^g \mathbf{x} \quad (2.9)$$

where the matrix  $\mathbf{K}^g$  involved the introduced APs:

$$\mathbf{K}^g = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.10)$$

As a conclusion, projective geometry allows linear representations for the well-known photogrammetric equations. Due to the characteristics of linearity, there is no need for the approximate values as required by the iterative solutions. Compared to the classical photogrammetric equations, this turns out to be a major advantage. However, some drawbacks are associated with these linear solutions (over-parameterization, non-linear image distortion is not considered, etc.). Therefore, the solutions for orientation obtained through linear equations will never be accurate enough based on a photogrammetric point of view, but provide an easy way of getting approximate values for the required unknown parameters (Ressl, 2004).

## CHAPTER 3

### PRE-PROCESSING AND 2D LINE SEGMENT EXTRACTION

It is a widely-accepted and well-known issue that the success and performance of image matching is highly affected by the radiometric characteristics of the aerial images utilized. The summary of previous work on line matching provided in Chapter 2 proves that, so far, only a limited number of studies have used and tested digital images for line matching. However, today, a large number of digital aerial cameras are available (see Chapter 2 section 2.1) and the information content of a digital image is proven to be much better than the information content of an analogue aerial image (see e.g., Jacobsen, 2007; Jacobsen, 2008). Therefore, utilizing digital images instead of analogue ones may reduce the radiometric instabilities and may lead significant improvements during the matching of the edge/line segments.

In this perspective, the general radiometric problems of aerial images can be summarized as (Zhang, 2005):

- Image noise is still a problem even in digital images while many feature extraction and image matching algorithms are sensitive to image noise. Therefore, the images have to be pre-processed in order to reduce the image noise.
- Poor image contrast, i.e. the peak of their histogram is typically towards the darker grey values, with the right part of the histogram decreasing smoothly towards the higher values.

- Radiometric problems caused by the variations in the sensor view angle, the sun angle and shadowing, the seasons and the atmospheric conditions should also be considered.

One important different aspect from the review of the line matching literature is that while aerial images have been rich of multispectral information, this fact was completely discarded or not efficiently used during the low level processing such as filtering, edge detection etc. In general, the multispectral information was effectively integrated during the matching stage (Bignone et al., 1996; Henricsson and Baltsavias; 1997; Henricsson, 1998; Moons et al., 1998; Scholze, 2000; Scholze et al., 2000; Zhang and Baltsavias, 2000; Leloglu, 2001). However, multispectral aerial images also provide opportunities for the extraction of line features that cannot be detected in the grayscale images (Scholze et al., 2000, Koschan and Abidi, 2005) due to several reasons, such as low contrast, accidental object alignments etc.

In this chapter, we introduce a new framework for the 2D extraction of line features from multispectral aerial images. Fig. 3.1 summarizes the proposed 2D line extraction framework. In the first step, in order to maximize the performance of the line detection, existing multispectral information in aerial images is fully utilized throughout the steps of pre-processing and edge detection. First, a multi-level non-linear color diffusion filter that involves the discontinuity information existing in different image bands is applied. Next, the method of color boosting is implemented to improve the color contrast between the adjacent objects in a scene. The edges are successfully extracted by integrating the edge information available in different multispectral bands of the aerial images. To accurately describe the straight edge segments, a principal component analysis (PCA) technique is adapted. Thereafter, the extracted segments are converted to their line counterparts using orthogonal regression. Finally, the uncertainty and attributes (geometric and radiometric) of each line are computed.

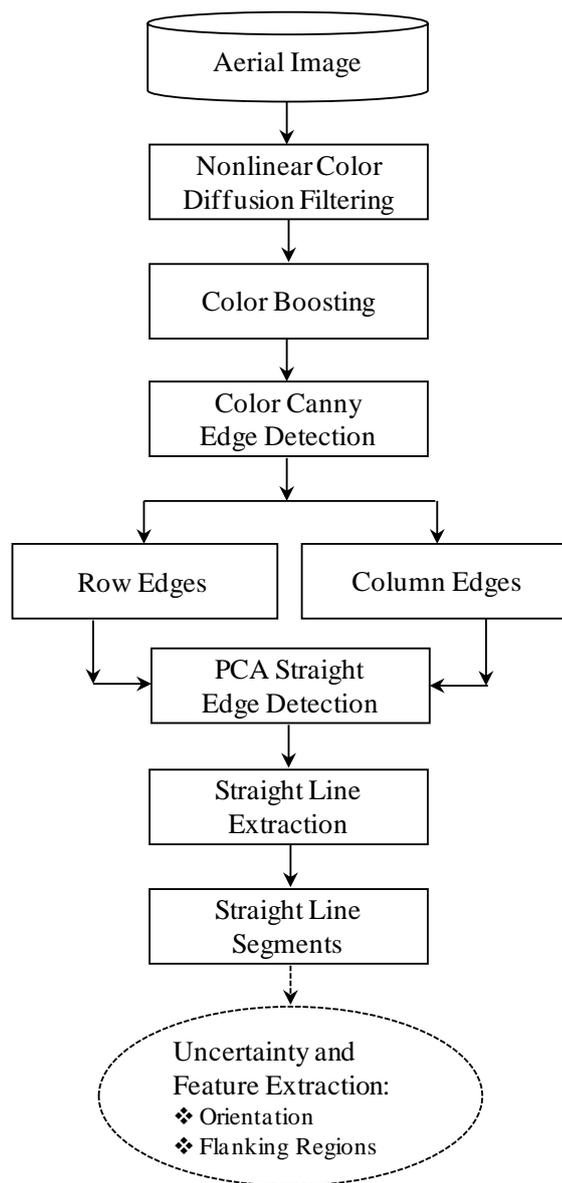


Figure 3.1 Flowchart of the proposed 2D line extraction approach.

### 3.1 Pre-processing

#### 3.1.1 Multi-Level Non-Linear Color Diffusion Filter

Image filtering for smoothing purposes can now be easily defined as one of the preliminary tasks for a wide variety of topics in the fields of photogrammetry and

computer vision. So far, a large number of smoothing algorithms have been developed (see e.g. Abramson and Schowengerdt, 1993; Weickert, 1997). In the edge/line matching context, most of the previous studies utilize widely accepted and well-known edge detectors such as Canny (Canny, 1986) or Canny-Derliche (Deriche, 1987). Those detectors utilize a classical Gaussian smoothing filter at an initial stage, and therefore, they can be regarded as homogeneous linear diffusion filters. Due to the Gaussian scale-space, those filters have two typical limitations (Weickert, 1997):

- Gaussian smoothing does not only reduce noise, but also blurs important features such as edges, and makes them harder to identify,
- Linear diffusion filtering results in dislocated edges when moving from finer to larger scales.

It is apparent that those limitations may deteriorate the final edge/line detection and matching performances. To improve the performance of the filtering, one attempt can be adaptively defining the size and/or shape of the Gaussian smoothing based on the underlying image structure (Nitzberg and Shiota, 1992) or to apply an inhomogeneous linear diffusion filtering process which further reduces smoothing around the edges detected by a fuzzy edge detector  $|\nabla f|$  (Weickert, 1997):

$$g(|\nabla f|^2) := \frac{1}{\sqrt{1+|\nabla f|^2/\lambda^2}} , \quad (3.1)$$

where  $g$  defines the diffusivity, and  $\lambda$  controls the sensitivity of diffusion to the detected edges. Although the diffusivity term in Eq. 3.1 is non-linear, the equation of diffusion for an inhomogeneous filtering stays linear as the original image is utilized for diffusivity in each time step  $t$  :

$$\delta_t u = \text{div}(g(|\nabla f|^2) \nabla u) \quad (3.2)$$

In Eq. 3.2,  $\delta_t u$  represents the concentration gradient,  $u$  is the actual image, and  $\text{div}$  is the divergence operator. Although the edges are better preserved than the homogeneous filtering, the fundamental problem of inhomogeneous filtering can be easily seen in Eq. 3.2, in which the diffusivity term will always be dependent on the differential structure of the original image. Thus, for large values of  $t$ , the filtered image may reveal some artifacts (Weickert, 1997). A solution to this problem can be achieved by using the gradient of the actual image  $u(x,t)$  instead of the original image  $f$  in the diffusivity term  $g$ . This leads to a non-linear diffusion equation (Perona and Malik, 1987):

$$\delta_t u = \text{div}(g(|\nabla u|^2) \nabla u) \quad (3.3)$$

The outcome after applying the non-linear diffusion filter is no or considerably reduced blurring around the edges. Although this seems to be a desired result for a general smoothing task, unsurprisingly, the noise around the edges cannot be removed completely since the blurring around the edges is entirely inhibited. A possible solution for this problem is to convert the scalar diffusion term  $g$  to a diffusion tensor leading to anisotropic diffusion filtering. However, a different and easy alternative way to solve the problem is to apply a multi-level non-linear diffusion filtering by jointly modifying the sigma parameter ( $\sigma$ ) and diffusion sensitivity ( $\lambda$ ) in each iteration level. The sigma parameter ( $\sigma$ ) is involved in the edge detection part in which the gradient is extracted. Therefore, decreasing the sigma parameter causes the gradient to be calculated at a less blurred image. Since the lambda parameter determines the gradient level to be diffused, an increase on the lambda parameter is required to reduce the noise level around the discontinuities. Thus, jointly modifying these two parameters enables us to lessen the noise around the image discontinuities. In this thesis, a three-level smoothing chain (decreasing the parameter sigma while increasing the parameter lambda) is designed to diminish the noise level around the discontinuities.

Up to this part, the smoothing approach works for grayscale images, since the edge detector  $|\nabla f|$  in Eq. 3.1 works on single band images. However, in our domain, we have aerial images that involve multi-band information. Therefore, identifying edges within a multi-band environment has crucial importance, since the smoothing is only restricted to the image parts where the discontinuities do not exist. In a multiband environment, for sure, the simplest idea would be to diffuse all available bands separately (Weickert, 1997). However, this may lead to different edge locations for each channel; thus, the final smoothed image may not be formed properly. One other basic approach is the calculation of the derivatives of each channel separately and adding them to produce a single combined gradient. However, for each channel, the derivatives of a color edge can be in opposing directions. Therefore, a summation of the derivatives per channel will discard the correlation between the color channels (Di Zenzo, 1986). In order to better locate the edges considering the available multi-band information, we adapted a gradient computed through tensor mathematics (see details in part 3.2.1) (Weijer et al., 2006b). By this way, edge information obtain from different bands strengthen each other to improve the performance of the multi-level non-linear filter. This can be done in an easy way; the grayscale edge detector  $|\nabla u|$  for each time step  $t$  is replaced by a color edge detector (see details in part 3.2.1) to perform the gradient extraction. Thus, in each step, the gradient of the actual image  $u(x,t)$  is computed by a multi-band edge detector, and subsequently utilized in the diffusion process.

In Fig. 3.2 and 3.3, two examples are given to illustrate the performance of the proposed multi-level non-linear color diffusion filter. The examples are intentionally selected from analogue images to further expose the filter performance. It is clear from those figures that, the dominant noise has been suppressed successfully from the images without blurring the edges. Besides, most of those critical edges, even the ambiguous ones (especially the edges belonging to dormers), are clearly improved and strengthened.



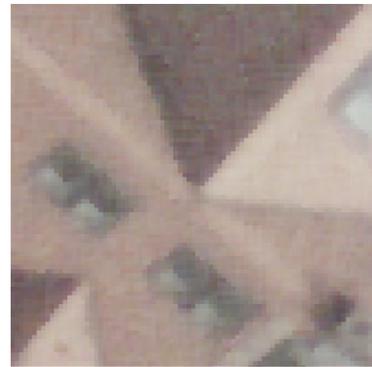
(a)



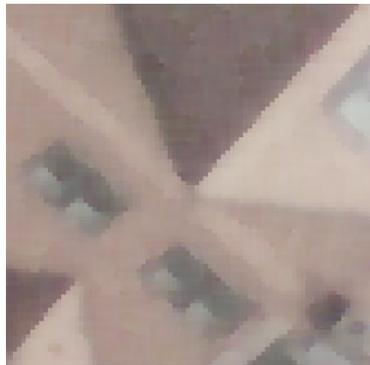
(b)



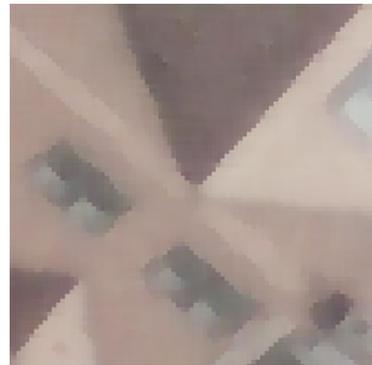
(c)



(d)



(e)



(f)

Figure 3.2 (a) Raw aerial image #1 and (b) the filtered result. (c-f) illustrates the evolution of the diffusion filtering process. (c) Original patch, (d) result of the first-level diffusion ( $\lambda = 0.03$ ,  $\sigma = 2$ ), (e) result of the second-level diffusion ( $\lambda = 0.05$ ,  $\sigma = 1$ ), and (f) result of the third-level diffusion ( $\lambda = 0.075$ ,  $\sigma = 0.5$ ).

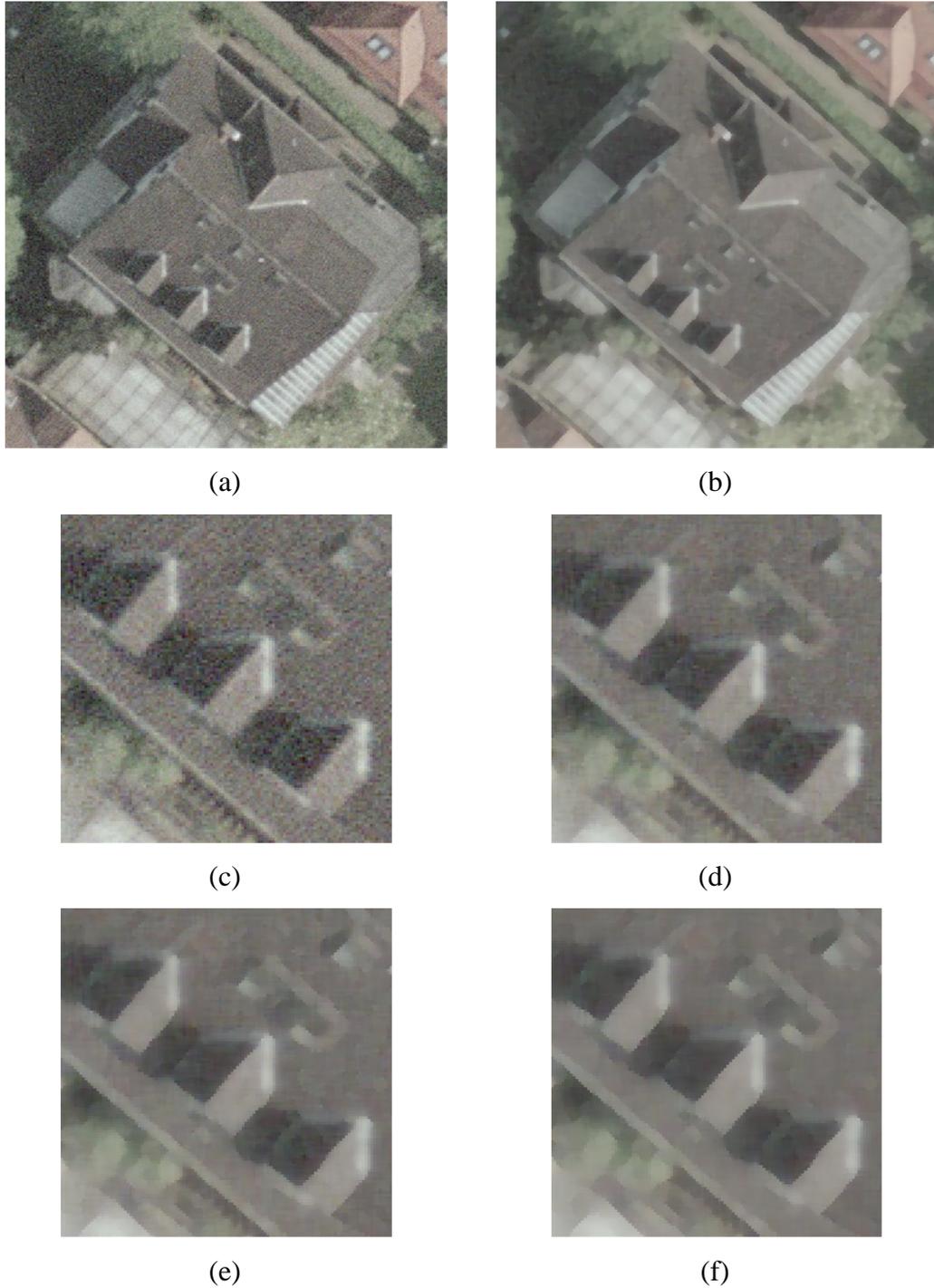


Figure 3.3 (a) Raw aerial image #2 and (b) the filtered result. (c-f) illustrates the evolution of the diffusion filtering process. (c) Original patch, (d) result of the first-level diffusion ( $\lambda = 0.03$ ,  $\sigma = 2$ ), (e) result of the second-level diffusion ( $\lambda = 0.05$ ,  $\sigma = 1$ ), and (f) result of the third-level diffusion ( $\lambda = 0.075$ ,  $\sigma = 0.5$ ).

### 3.1.2 Color Boosting

The goal of color boosting is to improve the apparent color difference between adjacent objects in a scene. For the aerial images (especially for analog cameras), the contrasts in the RGB values caused by the color variations are generally not high enough to exploit this distinction. Therefore, the idea is to amplify the color variations between the objects (for example, a building roof and its background) before the edge detection to find and extract the edges that cannot be detected due to low color variation. We utilized the boosting technique developed by Weijer et al. (2006a). First, the original RGB color space is transformed to the de-correlated Opponent Color Space ( $o_1$ ,  $o_2$ , and  $o_3$ ):

$$\begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \quad (3.4)$$

Next, to improve the color contrast in the images, color directions of the opponent space ( $o_1$  and  $o_2$ ) are selected and multiplied with a factor of  $k$  ( $k > 1$ ). Finally, the modified opponent color space is back-transformed to the boosted version of RGB color space:

$$\begin{pmatrix} R_b \\ G_b \\ B_b \end{pmatrix} = \begin{pmatrix} k \left( \frac{\sqrt{3}o_1 + o_2}{\sqrt{6}} \right) + \frac{o_3}{\sqrt{3}} \\ -k \left( \frac{\sqrt{3}o_1 + o_2}{\sqrt{6}} \right) + \frac{o_3}{\sqrt{3}} \\ -2k \frac{o_2}{\sqrt{6}} + \frac{o_3}{\sqrt{3}} \end{pmatrix} \quad (3.5)$$

The results of color boosting method based on various boosting factors are illustrated in Fig. 3.4 and 3.5. The boosted results prove that, for each small  $k$ , the color differences between adjacent objects in a scene are improved. However, on the contrary to the results provided by small boosting factors ( $k < 5$ ), larger  $k$



(a)



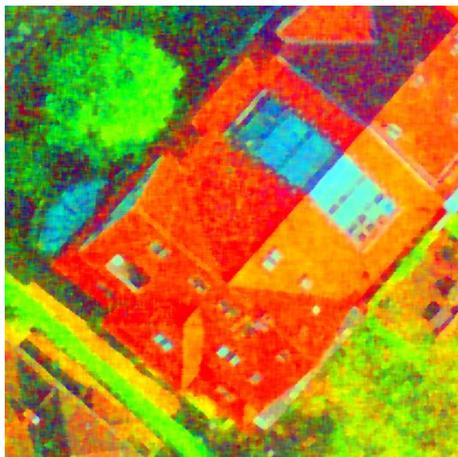
(b)



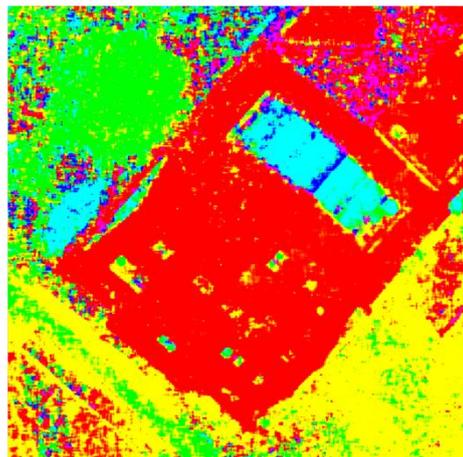
(c)



(d)



(e)



(f)

Figure 3.4 (a) Image #1 after discontinuity preserving filtering, and the results of color boosting with (b)  $k = 2$ , (c)  $k = 3$ , (d)  $k = 5$ , (e)  $k = 20$ , and (f)  $k = 200$ .



(a)



(b)



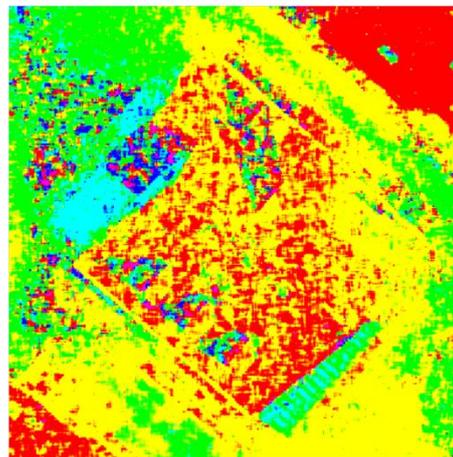
(c)



(d)



(e)



(f)

Figure 3.5 (a) Image #2 after discontinuity preserving filtering, and the results of color boosting with (b)  $k = 2$ , (c)  $k = 3$ , (d)  $k = 5$ , (e)  $k = 20$ , and (f)  $k = 200$ .

values starts to lose some of the critical edges that are formed by shading effects. Not surprisingly, for extremely large boosting factors (such as  $k = 200$ ), the boosting results converge to a stable constant boosted image which can be regarded as a very simple color classification output. For those cases, the boosted output of a three-band aerial image have at most 8 color classes (black, white, blue, green, red, cyan, magenta, and yellow) based on the dominant color value of each pixel in the input RGB image.

One important property of color boosting is that it is highly sensitive to the noise level inherent in aerial images, and if the noise in the images is not suppressed adequately, it also tends to boost the noise level of the images. Fig. 3.6 clarifies this issue. Color boosting with an excessive boosting factor ( $k = 200$ ) is applied to

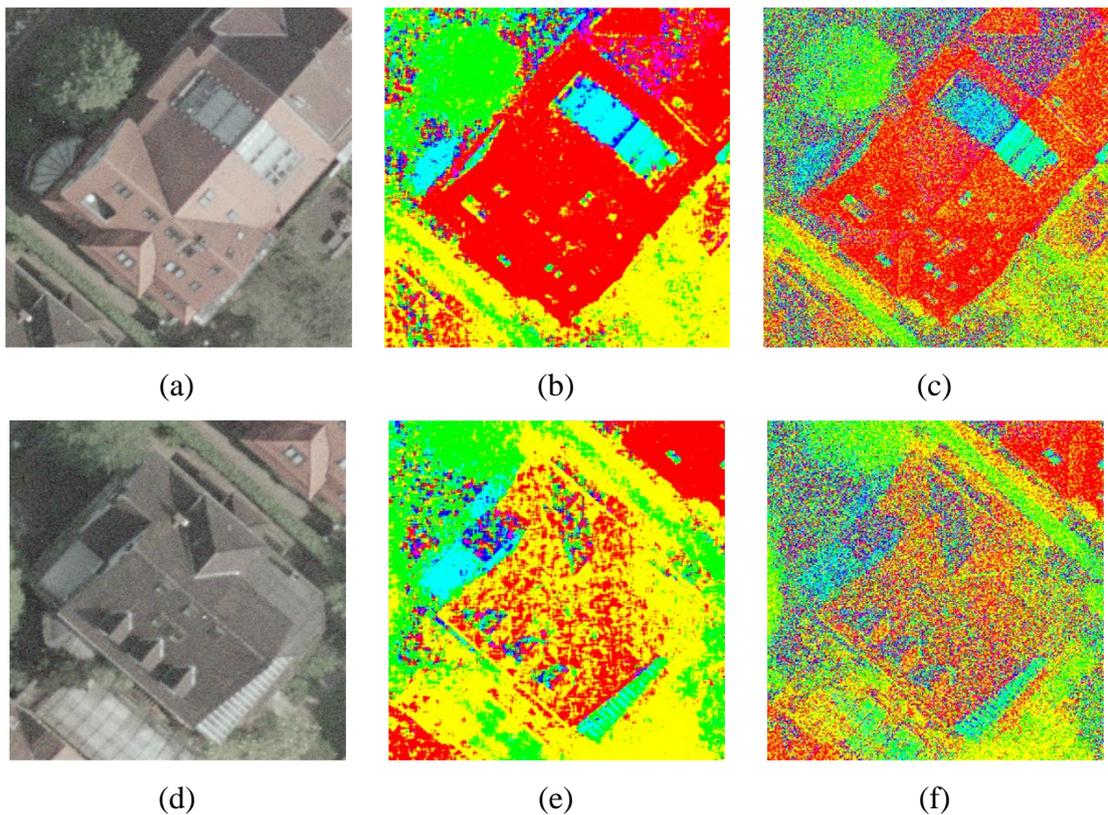


Figure 3.6 (a, d) The original images, boosted ( $k = 200$ ) images (b, e) after discontinuity preserving filtering, and (c, f) without performing any filtering operation.

the original and filtered images and the results prove that it is essential to apply color boosting to the noise filtered images.

## **3.2 Line Extraction**

### **3.2.1 Color Edge Detection:**

One way or another, the final performances of the developed methods mainly rely on the performance of a basic procedure known as the detection of the low level features, i.e. edges, in aerial images. Until now, a wide variety of edge detection techniques have been developed; however, the edge detectors used so far are mostly based on monochromatic techniques which either performs the detection on a single band, e.g. panchromatic, or a single band derived from several available bands using a certain method such as PCA. In the monochromatic context, a variety of edge detectors were proposed and/or used, some examples are Nevatia-Babu line finder algorithm (Huertas and Nevatia, 1988; Mohan and Nevatia, 1989), Suppression and Enhancement operator (Henricsson, 1998), Boldt hierarchical algorithm (Collins et al., 1998; Jaynes et al., 2003), Optimal Zero Crossing operator (Sahar and Krupnik, 1999), Prewitt operator (Turker and San, 2004) and Laplacian operator (You and Shiqiang, 2006). Besides those detectors, the Canny (Canny, 1986) or Canny-Deriche (Deriche, 1987) edge detectors were probably the most preferred ones (see e.g., Lin and Nevatia, 1998; Kim and Nevatia, 1999; Stassopoulou and Caelli, 2000; Cord et al., 2001; Bilen, 2004; Güler, 2004; Koc San, 2009; Habib et al., 2010). However, the low level object extraction performances of those detectors are always limited since they work on a single band or component. A very clear and good example that shows the potential of a multi-band (or color) processing is illustrated in Fig. 3.7. In the example, a color variant (Koschan and Abidi, 2005) and a classical grayscale Canny edge detector were applied to a checkerboard pattern consisting of three different color squares that have similar intensity values. The edge detection results in Fig. 3.7 unquestionably prove the importance of multi-band or color

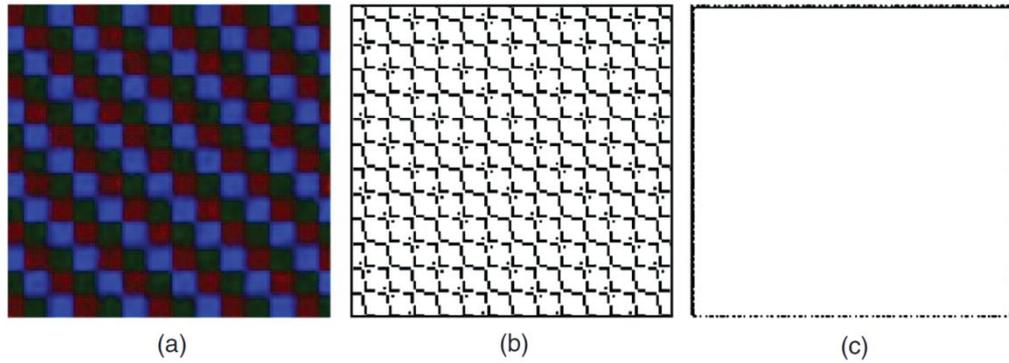


Figure 3.7 (a) Original image in a grid pattern, (b) result of a color variant canny edge detector, and (c) result of a grayscale canny edge detector (Koschan and Abidi, 2005).

edge processing. In this thesis, to maximize the performance of the edge detection, the multiband edge detection approach proposed by Weijer et al. (2006b) is utilized. They proposed a color Canny edge detection algorithm to accurately locate the edges in multispectral images. The algorithm mainly consists of calculation of the spatial derivatives of different image channels, and the computed derivatives are combined using tensor mathematics. In this way, differential structures of the bands in multispectral images are mutually supported, and an edge detection of better completeness is accomplished. Here we briefly review the approach; more details can be found in Weijer et al. (2006b).

Adding the differential structure of multispectral channels may cause cancellation even when apparent edges exist in the image (Di Zenzo, 1986). Thus, instead of dealing with edge directions defined in the range of  $[0, 2\pi]$ , it is more appropriate to work on the orientation ranges between  $[0, \pi]$ . By this way, during the summation of edge information, opposite directions will contribute and reinforce each other. Since the tensor of a vector and its 180 rotated counterpart vector are equal, tensors are convenient to describe color derivative vectors. Given a

grayscale image  $f$ , the structure tensor is given by (Di Zenzo, 1986; Förstner, 1994):

$$G = \begin{pmatrix} \overline{f_x^2} & \overline{f_x f_y} \\ \overline{f_y f_x} & \overline{f_y^2} \end{pmatrix}, \quad (3.6)$$

where the subscripts indicate spatial derivatives and the bar ( $\bar{\phantom{x}}$ ) indicates convolution with a Gaussian filter. For a multispectral image  $\mathbf{f} = (f^1, f^2, \dots, f^n)^T$ , the structure tensor can be defined as (Weijer et al., 2006b)

$$G = \begin{pmatrix} \overline{\mathbf{f}_x \cdot \mathbf{f}_x} & \overline{\mathbf{f}_x \cdot \mathbf{f}_y} \\ \overline{\mathbf{f}_y \cdot \mathbf{f}_x} & \overline{\mathbf{f}_y \cdot \mathbf{f}_y} \end{pmatrix}. \quad (3.7)$$

With the aid of the spatial derivatives, two eigenvalues of the tensor  $G$  in eq. 3.7 can be defined as

$$\lambda_1 = \frac{1}{2} \left( \overline{\mathbf{f}_x \cdot \mathbf{f}_x} + \overline{\mathbf{f}_y \cdot \mathbf{f}_y} + \sqrt{(\overline{\mathbf{f}_x \cdot \mathbf{f}_x} - \overline{\mathbf{f}_y \cdot \mathbf{f}_y})^2 + (2\overline{\mathbf{f}_x \cdot \mathbf{f}_y})^2} \right) \quad (3.8)$$

$$\lambda_2 = \frac{1}{2} \left( \overline{\mathbf{f}_x \cdot \mathbf{f}_x} + \overline{\mathbf{f}_y \cdot \mathbf{f}_y} - \sqrt{(\overline{\mathbf{f}_x \cdot \mathbf{f}_x} - \overline{\mathbf{f}_y \cdot \mathbf{f}_y})^2 + (2\overline{\mathbf{f}_x \cdot \mathbf{f}_y})^2} \right).$$

The most prominent local orientation can also be defined as (Weijer et al., 2006b)

$$\theta = \frac{1}{2} \arctan \left( \frac{2\overline{\mathbf{f}_x \cdot \mathbf{f}_y}}{\overline{\mathbf{f}_x \cdot \mathbf{f}_x} - \overline{\mathbf{f}_y \cdot \mathbf{f}_y}} \right). \quad (3.9)$$

By means of the large eigenvalue ( $\lambda_1$ ) in Eq. 3.8 and the orientation ( $\theta$ ) in Eq. 3.9, the color Canny edge detector can be introduced (Gevers, 2006):

- Compute the spatial derivatives,  $\mathbf{f}_x$  and  $\mathbf{f}_y$

- Compute eigenvalue ( $\lambda_1$ ) and its orientation ( $\theta$ )
- Apply non-maximum suppression on  $\lambda_1$  in the prominent direction.

Here, two minor adaptations enhance the results of the algorithm: (i) the output of the final gradient map is scaled between zero-and-one before further processing, which significantly reduces the remaining noisy edges, and (ii) a two level hysteresis thresholding is designed to have a better control on the final edge contours.

The results of the color edge detection are illustrated in Fig. 3.8. It can be seen that the edges describing the object boundaries are successfully extracted. Furthermore, the building boundaries that show only a very slight color difference are detected.

### **3.2.2 Straight Line Extraction:**

One of the common problems of image processing is to extract application specific geometric primitives (straight line segments, circles etc.) from the complete edge maps. In the domain of straight edges/lines, until now, a significant number of methods/detectors were proposed and some examples are the Nevatia-Babu detector (Nevatia and Babu, 1980), Burns edge detector (Burns, 1986), Hough Transform (Hough, 1962) and its variants (Ballard, 1981; Li et al., 1986; Illingworth and Kittler, 1987; Ben-Tzvi and Sandler, 1990; Princen et al. 1990), and Steger Line Extractor (Steger, 1998). On the other hand, more recently, eigenvalue (Guru et al., 2004) and PCA (Nagabhushan et al., 2005; Shekar et al., 2006; Lee et al., 2006) based approaches gained more attention from the community.

In this part, a two stage solution for the straight line extraction problem is proposed, (i) the extraction of straight edge segments, and (ii) fitting line



from aerial images. Furthermore, in this study, several drawbacks of the approach are identified and corrected.

The approach starts with a labeling procedure in which the extracted edges are classified into two distinct edge classes, row and column segments. To do that, four types of marks (row, column, cross, and single) are used, and the row and column edge segments are labeled using 8-neighbour connectivity. Thereafter, PCA method is used to extract two principal components of each row and column segment. To do that, the scatter matrices of edge segments must be computed (Lee et al., 2006):

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \quad (3.10)$$

If  $n$  is the number of pixels of a labeled segment and  $(x_i, y_i)$  is the coordinates of the  $i$ th pixel of that segment, the components of the scatter matrix  $S$  is calculated as (Lee et al., 2006):

$$\begin{aligned} s_{11} &= \frac{1}{n} \sum_{i=1, \dots, n} (x_i - x_m)^2, \\ s_{11} = s_{12} &= \frac{1}{n} \sum_{i=1, \dots, n} (x_i - x_m)(y_i - y_m), \\ s_{22} &= \frac{1}{n} \sum_{i=1, \dots, n} (y_i - y_m)^2, \end{aligned} \quad (3.11)$$

where  $x_m$  and  $y_m$  are the mean coordinate values of the pixels that fall in that segment. The first ( $\lambda_1$ ) and second ( $\lambda_2$ ) eigenvalues from the given scatter matrix in Eq. 3.10 can be computed as (Lee et al., 2006):

$$\lambda_1 = \frac{1}{2} \left( s_{11} + s_{22} + \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \right) \quad (3.12)$$

$$\lambda_2 = \frac{1}{2} \left( s_{11} + s_{22} - \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \right)$$

In principle, the straightness of an edge segment can be determined from the magnitude of the second eigenvalue ( $\lambda_2$ ) computed in Eq. 3.12. Thus, an absolute threshold ( $t_a$ ) can be forced on the second eigenvalues of each segment to determine whether the segment is straight or not. However, depending on the length of the edge segment, noise which arises from several short primitives may distort the computed straightness values of the segment. Due to this reason, the straightness of long segments may not be correctly determined with a single absolute threshold. Therefore, to solve this problem, Lee et al. (2006) proposed a relative thresholding scheme. The edge segments are labeled as straight if the eigenvalue of the second principal component ( $\lambda_2$ ) is less than a predefined value ( $t_i$ ) which is also a function relative to the length of the segment ( $l_i$ ):

$$t_i = \left( \frac{l_i}{l_{min}} \right)^2 t_a , \quad (3.13)$$

where  $l_{min}$  denotes the total number of pixels of the shortest line permitted. By this way, depending on the length of each segment, threshold values are automatically determined and used to test the straightness of each segment.

Although the method has proven to be fast and efficient, several problems are identified during the extraction of the straight edge segments. First, the input binary edge images are assumed to be segments that are only a single pixel wide. However, this is generally not the case for the output of the binary images generated by the color canny edge detection. Although non-maximum suppression is applied after the detection stage, this does not always guarantee one pixel wide edges extracted from color images, since separate spatial derivatives of the image bands are combined during edge detection. To solve this problem, we utilized the image skeleton technique (MatLab, 2009) to remove the redundant boundary

pixels of the binary edges. The technique ensures that the binary objects shrink to a minimally connected structure without breaking apart.

A different shortcoming observed is that, since the approach mainly depends on row and column primitives, after the edge detection stage, a straight object boundary in an image may be represented by more than one primitive. This is actually an expected output if the straight object boundary actually occurs in the same arrangement in the image space. However, in some cases, accidental edge alignments may cause the straight edges to break apart into several straight pieces which in turn significantly reduce the performance of the straight edge detector. Several examples are illustrated in Fig 3.9a. The figure points out some critical edge locations in which the straight edges are split into row and column segments. Therefore, those cases must be handled and corrected before the determination of

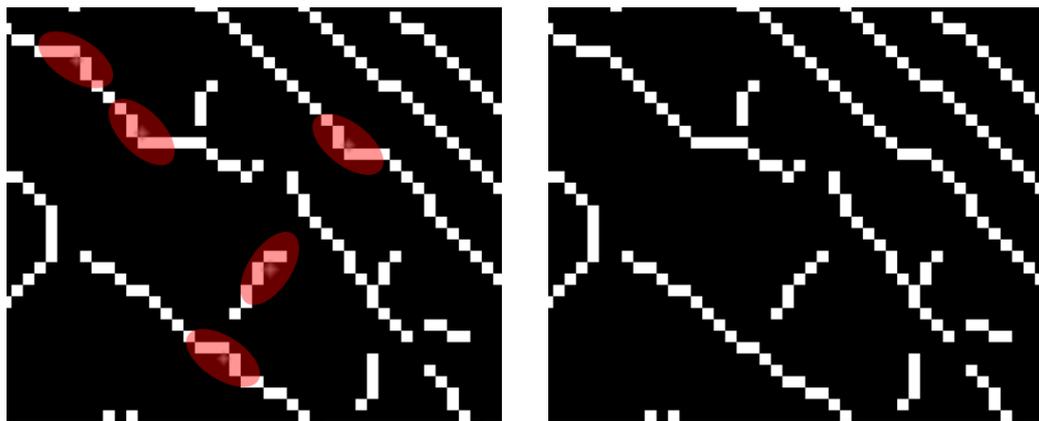


Figure 3.9 (a) Artificial critical locations and (b) the corrected result.

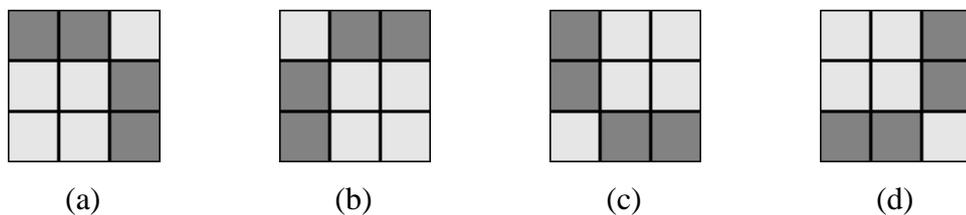


Figure 3.10 Masks utilized to identify and correct the artificial critical locations.

row and column primitives. In this thesis, four different masks (Fig. 3.10) are utilized to identify those critical configurations. Thereafter, those locations are corrected by replacing them with the cross primitives (Fig. 3.9b).

Another critical shortcoming occurs if two same label (for example two row-directional) edge segments are connected with a cross primitive junction, in this case the algorithm is not capable to determine the correct straightness value. Unfortunately, this type of line to line combinations is not rare in aerial images. To solve the problem, four different directional masks with 5x5 pixels (Fig. 3.12) are applied to identify problematic intersections. An example is given in Fig. 3.11. In the figure, two reddish ellipses cover only a single segment composed of only row primitives. Thus, it is impossible to validate the segment as a single straight row edge, as the second-eigenvalue ( $\lambda_2$ ) of this segment may simply exceed the

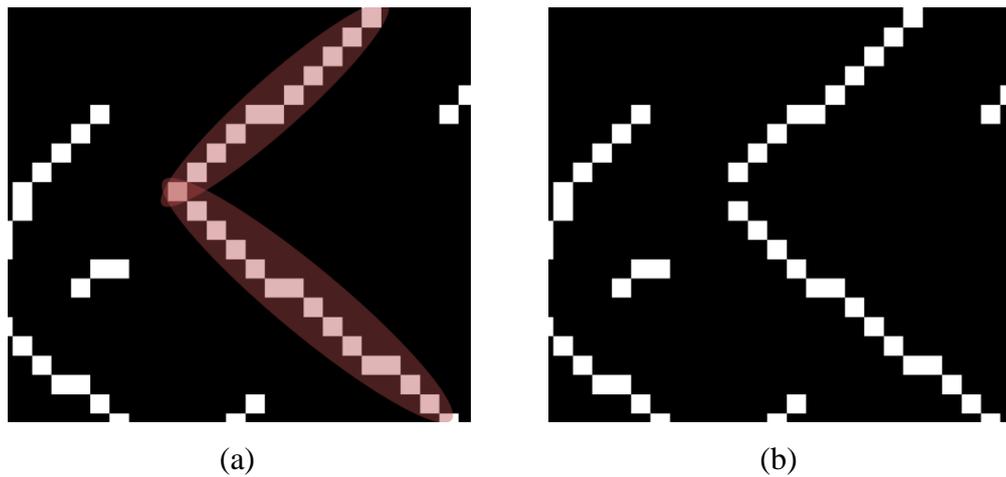


Figure 3.11 (a) A problematic same-label intersection and (b) the corrected result.

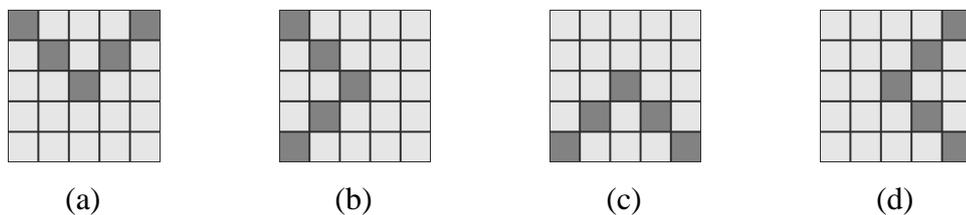


Figure 3.12 Masks utilized to identify and correct the critical intersections.

predefined straightness threshold  $t_i$  in Eq. 3.13. Thus, we identify the intersection point of this segment and thereafter, the problematic intersection is removed prior to the computation of the second-eigenvalue ( $\lambda_2$ ).

The results of the straight edge detector based on the two image patches are illustrated on Fig. 3.13 and 3.14. The row and column edges are extracted from the edge images generated by the color Canny edge detector (Fig. 3.13c and 3.14c). After that, the edge segments are labeled as row and column primitives, and the straightness of each segment is tested based on the second-eigenvalue ( $\lambda_2$ ) of the PCA analysis. In order to better visualize the difference between the absolute and relative thresholding, same thresholds ( $t_a = 0.1$ ) are applied to both image patches. As seen in Fig. 3.13f and 3.14f, most of the straight edge segments were missed with the absolute eigenvalue threshold of 0.1. However, the detected straight segments can be considered as the most reliable ones. On the other hand, more straight segments are observed in Fig. 3.13i and 3.14i, in which the thresholding is performed in a relative manner with same eigenvalue threshold. In that case, the effects are taken into account for longer edge segments; thus, a large number of straight segments are detected. Nevertheless, it is also clear from those results that, although the relative thresholding is applied during the selection of straight edge segments, several important object boundaries are still missing. Thus, the noise affects that may cause curvilinear straight object boundaries must be considered, and handled with a larger eigenvalue threshold. In Fig. 3.13m and 3.14m, straight edge segments that are extracted with a large eigenvalue threshold ( $t_a = 1.6$ ) is shown. In that case, we observe that almost all of the straight segments are successfully located. Not surprisingly, the large threshold also recognizes and labels some short curved edge segments as straight objects. However, those segments can be treated and handled in the next step, in which the straight edge segments are converted into straight line segments.

In this study, we refer to a line segment, a single straight object that is composed of only two endpoints ( $x_s, y_s; x_f, y_f$ ). To accurately describe the line segments, we

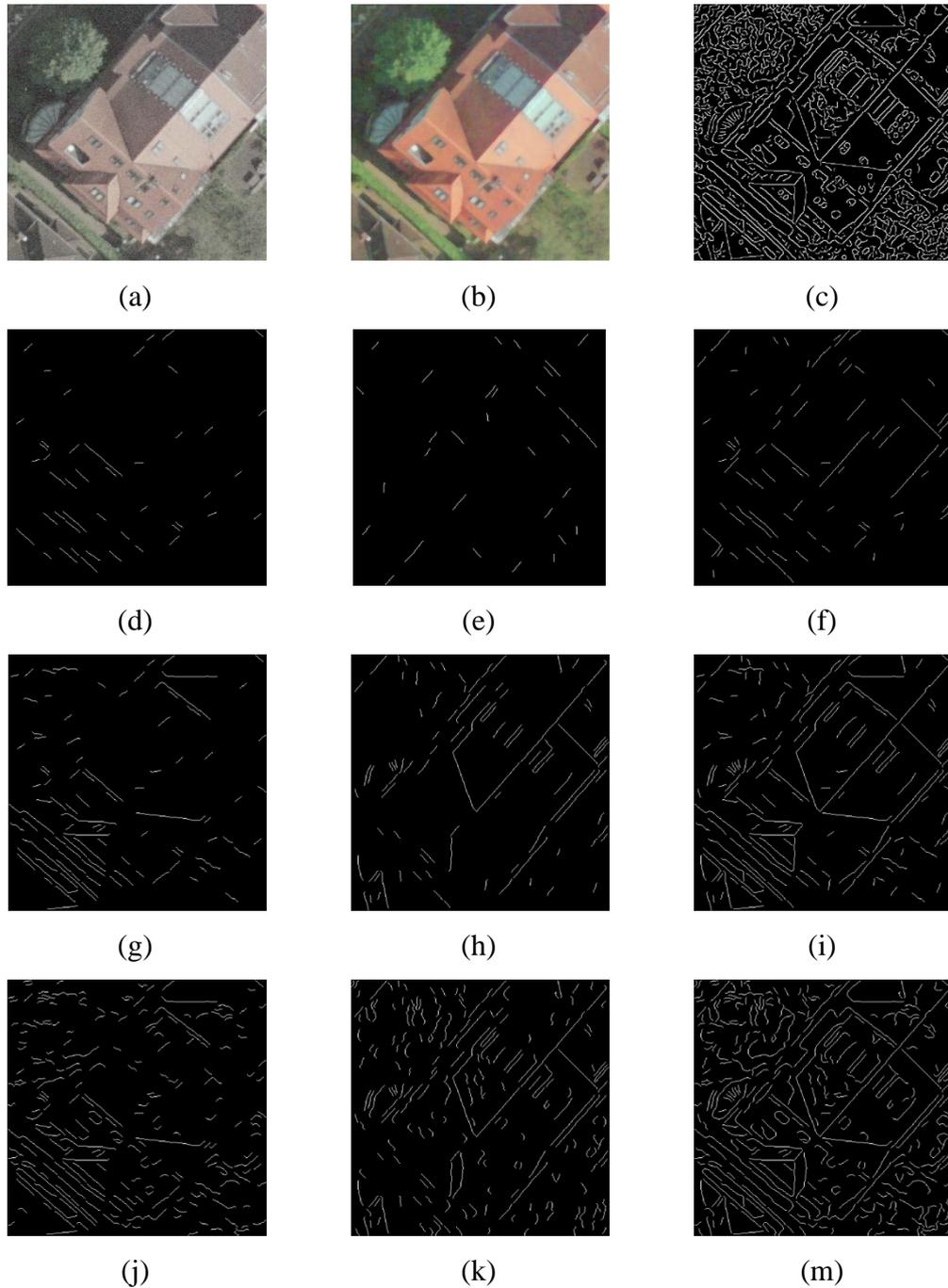


Figure 3.13 (a) Original image #1, (b) filtered and boosted image, and (c) color Canny edges. Straight row, straight column and final straight edges are illustrated in first, second and third columns, respectively. (d-f) Straight edges with absolute thresholding with  $t_a = 0.1$ , (g-i) straight edges with relative thresholding with  $t_a = 0.1$ , and (j-m) straight edges with relative thresholding with  $t_a = 1.6$ . For all cases, minimum segment length is selected as 10 pixels.

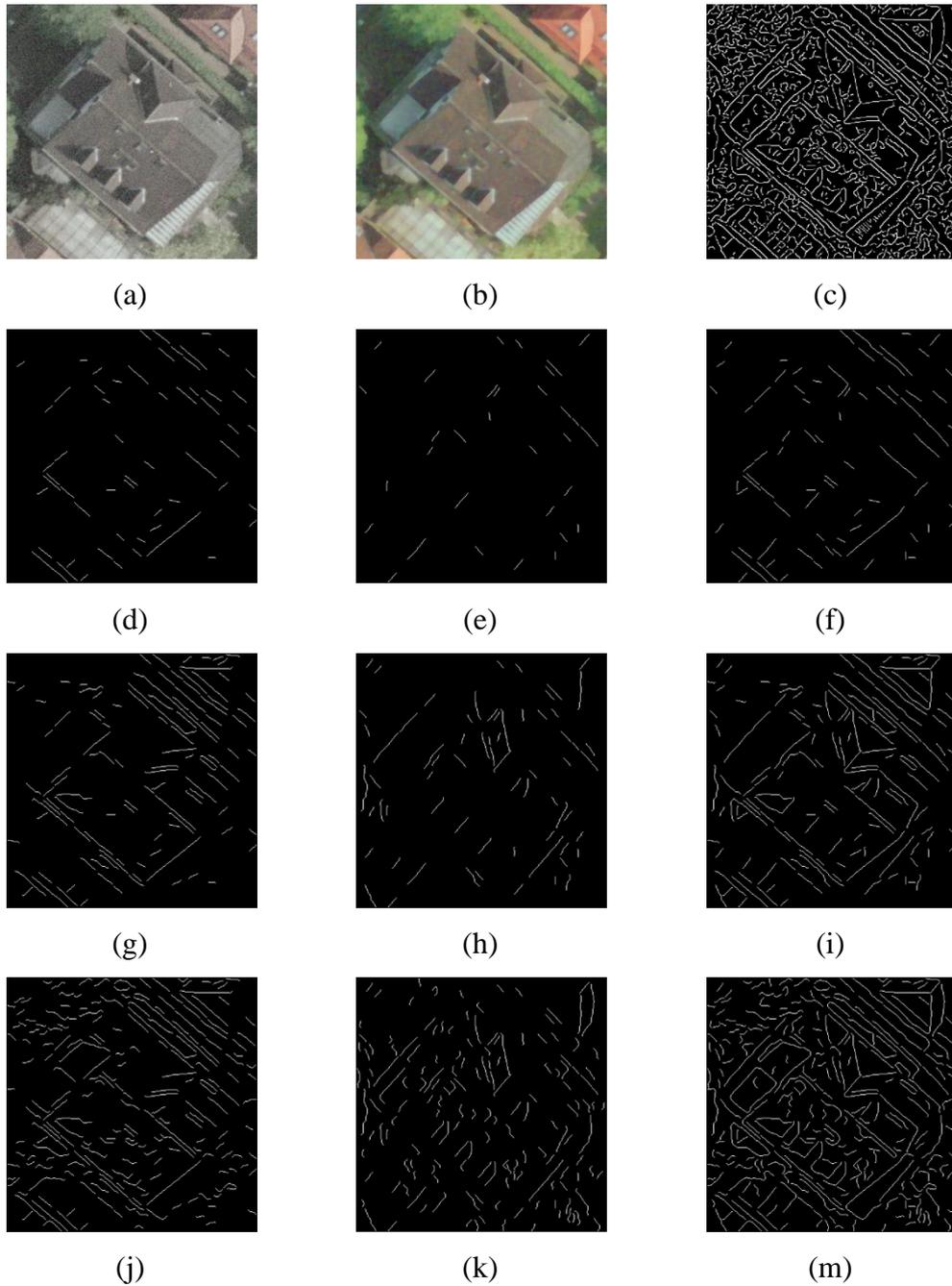


Figure 3.14 (a) Original image #2, (b) filtered and boosted image, and (c) color Canny edges. Straight row, straight column and final straight edges are illustrated in first, second and third columns, respectively. (d-f) Straight edges with absolute thresholding with  $t_a = 0.1$ , (g-i) straight edges with relative thresholding with  $t_a = 0.1$ , and (j-m) straight edges with relative thresholding with  $t_a = 1.6$ . For all cases, minimum segment length is selected as 10 pixels.

consider that the accumulation of coordinates  $(x, y)$  of all pixels forming a straight edge segment be a set of two-dimensional data points  $(x_i, y_i)$ ,  $i = 1 \dots n$ . Thus, the aim is to fit the line segments to those set of points  $(x_i, y_i)$  in which both  $x_i$  and  $y_i$  are assumed to be influenced by an isotropic noise model that has same variances for each coordinate:

$$\Sigma = \begin{bmatrix} \sigma_{pixel}^2 & 0 \\ 0 & \sigma_{pixel}^2 \end{bmatrix}, \quad (3.14)$$

where  $\Sigma$  is the covariance matrix associated with each pixel and  $\sigma_{pixel}$  is the standard deviation of the edge pixel coordinate noise distribution (Madsen and Christensen, 1995). At this point, a line representation is required (Fig. 3.15); in this part, the line representation proposed by Deriche et al. (1991) is utilized:

$$x \sin(\tau) - y \cos(\tau) + \rho = 0 \quad (3.15)$$

The required parameters  $(\tau, \rho)$  of the straight line in Eq. 3.15 can be computed by

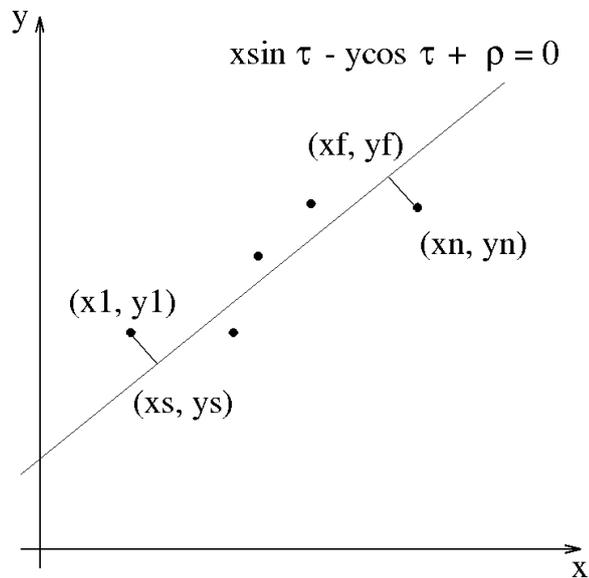


Figure 3.15 Fitting a straight line to a set of edge pixels (Madsen and Christensen, 1995)

$$\tau = \frac{1}{2} \arctan\left(\frac{b}{a-c}\right) \quad (3.16)$$

$$\rho = \bar{y} \cos(\tau) - \bar{x} \sin(\tau) \quad (3.17)$$

where

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i & \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \\ a &= \sum_{i=1}^n (x_i - \bar{x})^2 & b &= 2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) & c &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

After fitting the infinite straight line, the endpoints of a line segment can be defined by projecting the extent of edge segment onto the infinite line (Madsen and Christensen, 1995):

$$\begin{aligned} x_s &= x_1 - (x_1 \sin(\tau) - y_1 \cos(\tau) + \rho) \sin(\tau) \\ y_s &= y_1 + (x_1 \sin(\tau) - y_1 \cos(\tau) + \rho) \cos(\tau) \end{aligned} \quad (3.18)$$

$$\begin{aligned} x_f &= x_n - (x_n \sin(\tau) - y_n \cos(\tau) + \rho) \sin(\tau) \\ y_f &= y_n + (x_n \sin(\tau) - y_n \cos(\tau) + \rho) \cos(\tau) \end{aligned}$$

Assuming the basic edge pixel noise model presented in Eq. 3.14, it can be shown that the covariance matrix of the parameters of the infinite line,  $(\tau, \rho)$  becomes (Madsen and Christensen, 1995):

$$\Lambda = \frac{(a+c)\sigma_{pixel}^2}{(a-c)^2+b^2} \begin{bmatrix} 1 & -d \\ -d & d^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{\sigma_{pixel}^2}{n} \end{bmatrix} \quad (3.19)$$

$$d = \bar{y} \sin(\tau) - \bar{x} \cos(\tau) \quad (3.20)$$

Once again, the coordinate covariance matrices of the two segment endpoints  $(x_s, y_s; x_f, y_f)$  can also be computed from Eq. 3.18. Using the starting endpoint as example, the vector  $\vec{v}_s = [x_s \ y_s]$  expresses the endpoint coordinates as

functions of the coordinates of a vector  $\vec{p}_1 = [\tau \ \rho \ x_1 \ y_1]$ . If  $\Delta^s$  defines the covariance matrix associated with the coordinates of the endpoint:

$$\Delta^s = \begin{bmatrix} \sigma_{x_s x_s}^2 & \sigma_{x_s y_s}^2 \\ \sigma_{x_s y_s}^2 & \sigma_{y_s y_s}^2 \end{bmatrix} = \frac{\partial \vec{v}_s}{\partial \vec{p}_1} \Gamma \frac{\partial \vec{v}_s}{\partial \vec{p}_1}^T \quad (3.21)$$

where  $\partial \vec{v}_s / \partial \vec{p}_1$  is the 2x4 Jacobian matrix and  $\Gamma$  is the 4x4 covariance of  $\vec{p}_1$ :

$$\Gamma = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Sigma \end{bmatrix} \quad (3.22)$$

Thus,  $\Delta^s$  is a 2x2 covariance matrix expressing the two variances of the 2D distribution function of endpoint coordinates (Madsen and Christensen, 1995).

In some cases, a single straight edge segment may be formed by more than one straight line segment. In this case, a recursive strategy is utilized. First, an infinite line model is estimated from the entire set. Next, the edge points that support the line model (based on a specific orthogonal distance tolerance) are removed from that straight edge segment. The estimation is repeated with the remainder of the points to find the next infinite line model. The iteration stops if the number of pixels left is less than the number of pixels of the shortest line permitted ( $l_{\min}$ ) or there are no pixels left on the straight edge segment. Fig. 3.16b and 3.16e illustrate the line segments that are estimated from the detected straight edge segments. The overlay figures with the original images in 3.16c and 3.16f indicate that almost all of the straight line boundaries that are related with objects and their details are well extracted and most of the short curved segments are eliminated.

After the extraction of line segments, several geometric attributes of each line segment, such as orientation ( $\theta_l$ ) and length ( $L_l$ ), can be computed. The orientation attribute ( $\theta_l$ ) of each line can be easily computed with the following equation:

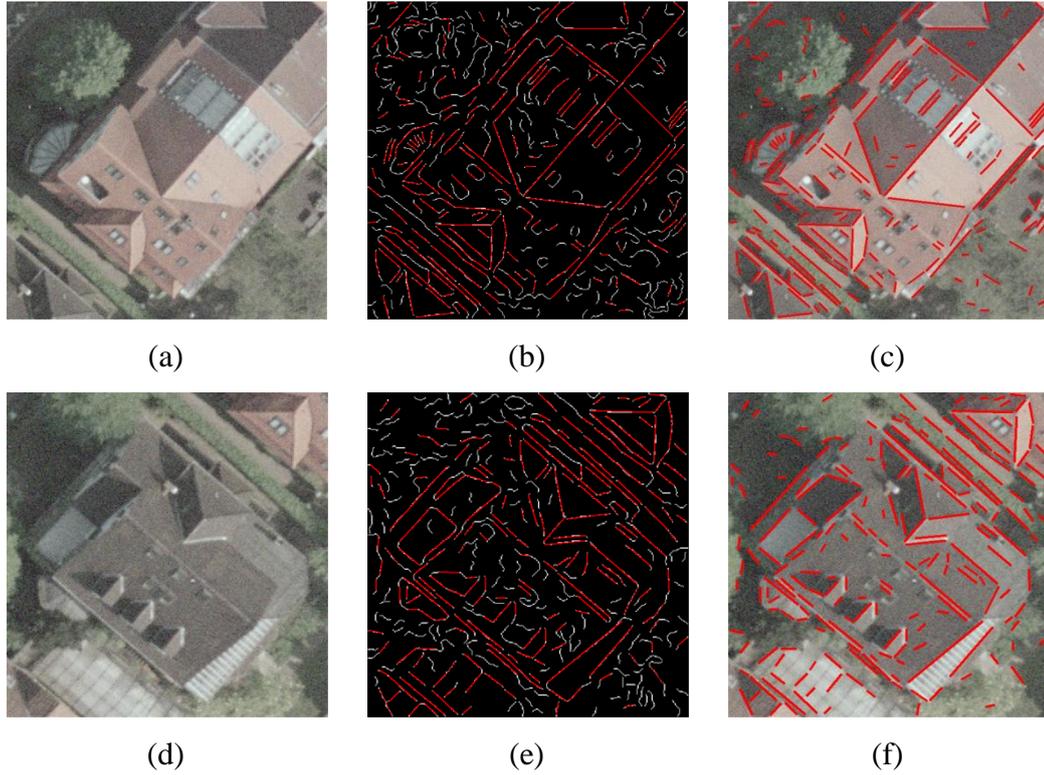


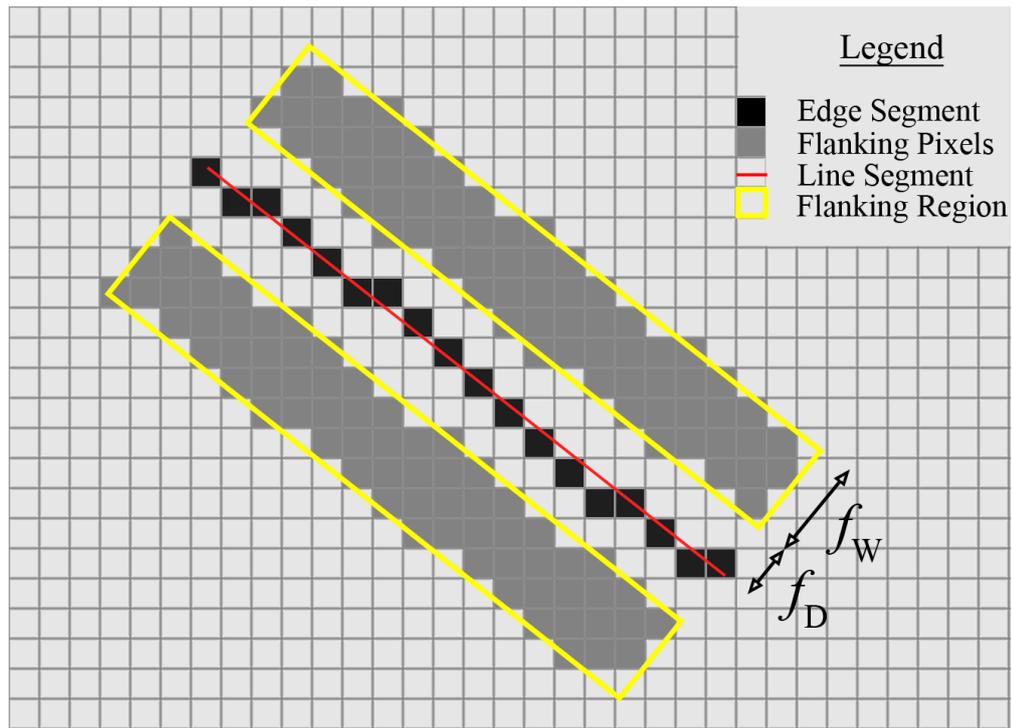
Figure 3.16 Original images (a, d), straight line segments overlaid to straight edge segments (b, e), and straight line segments overlaid to original images (c, f).

$$\theta_l = \text{mod} \left\{ \arctan \left( \frac{x_s - x_f}{y_s - y_f} \right) + \pi, \pi \right\}, \quad (3.23)$$

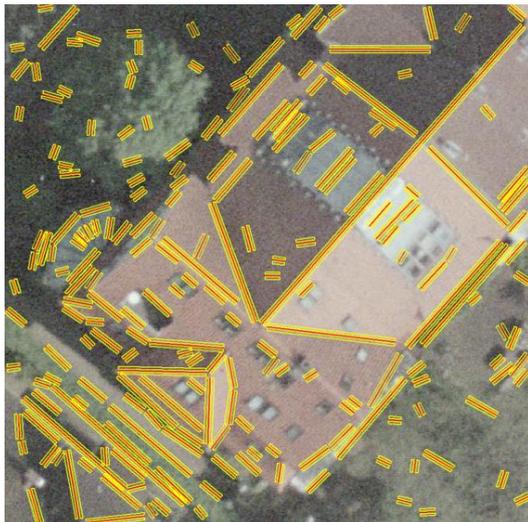
where  $\text{mod}\{a, b\}$  is the modulus operator that guarantees the computed angle value is bounded with  $0-\pi$  radians. The computation of the length attribute of each line is also straightforward:

$$L_l = \sqrt{(x_s - x_f)^2 + (y_s - y_f)^2} \quad (3.24)$$

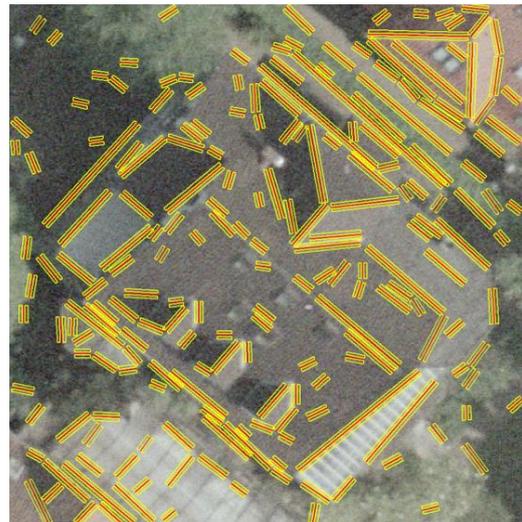
Radiometric attributes which obtain the image surface reflectance values in the vicinity of each line segment can be computed by generating flanking regions on each side of the line segments (Fig. 3.17). Two parameters, flanking distance ( $f_D$ )



(a)



(b)



(c)

Fig. 3.17 (a) An edge segment and the left and right flanking regions. (b, c) The generated flanking regions for each line segment.

and flanking width ( $f_w$ ), are sufficient to determine the flanking regions on each side (Fig. 3.17a). Once the flanking regions are generated, the pixels that fall in each region must be determined. However, this issue is not trivial, since some of

the pixels on the edge of a region are only partially covered (Fig. 3.17a). At this point, a simple rule is sufficient to decide which pixels will be included to the region: if a pixel's central point is inside the boundary of a flanking region, the pixel is regarded as inside the flanking region. Once the flanking pixels are determined, the radiometric values available in each band are collected. However, since a flanking region may also contain irrelevant objects or disturbances, e.g., chimneys and shadows, with spectral characteristics other than the main reflectance, these disturbances (outliers) must be taken into account during the computation of the radiometric attributes (Henricsson, 1998). In this thesis, the mean values of each band in the flanking region are robustly estimated using the method of minimum covariance determinant (Meucci, 2005). The method searches for an observation  $\mathbf{x}_t$  such that if we remove it from a given set of observations  $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ , the determinant of the resulting sample covariance is reduced the most. This would mean, after removing that observation from the set of observations, the sample mean and covariance shrinks the most, and thus that observation ( $\mathbf{x}_t$ ) is the farthest outlier in the sample. To perform this, the following equation must hold (Meucci, 2005):

$$|\mathbf{U}_{(-t)}^T \mathbf{U}_{(-t)}| = (1 - \lambda_t) |\mathbf{U}^T \mathbf{U}|, \quad (3.25)$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{x}_1^T - \hat{\boldsymbol{\mu}}^T \\ \vdots \\ \mathbf{x}_T^T - \hat{\boldsymbol{\mu}}^T \end{bmatrix}_{TxN}$$

In Eq. 3.25,  $\hat{\boldsymbol{\mu}}^T$  is the sample mean of the data,  $\mathbf{U}_{(-t)}$  denotes the matrix after removing the  $t$ -th row,  $N$  is the number of bands, and  $\lambda_t$  is the  $t$ -th element of the diagonal of the following matrix:

$$\lambda_t \equiv (\mathbf{U}(\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T)_{tt} \quad (3.26)$$

It can be proved that  $0 \leq \lambda_t \leq 1$ . Thus, the farthest outlier corresponds to the highest value of  $\lambda_t$ , unless  $\lambda_t = 1$ : in that case, if we remove the  $t$ -th observation, the sample covariance in Eq. 3.25 turns out to be singular (Meucci, 2005). In this way, in each iteration, the minimum covariance determinant method removes a single observation from the entire set which has the highest value of  $\lambda_t$ . However, since we do not know a priori the total number of outliers in each flanking region, we defined a breakdown factor of 50%. Thus, we stop the iterations if the number of observations in the dataset is less than half of the original number in the flanking regions.

## CHAPTER 4

### STEREO LINE MATCHING

Once the straight line segments are independently extracted for both of the stereo images, a matching strategy is required to find the line correspondences between the images. In this chapter, we propose a new line matching approach to establish the correspondences in a stereo manner (Fig. 4.1). The proposed approach initially generates reference line pairs in the first image and collects all potential matching candidate pairs from the second image with the aid of a-priori known image to image geometry. Next, the number of candidate pairs for the matching is

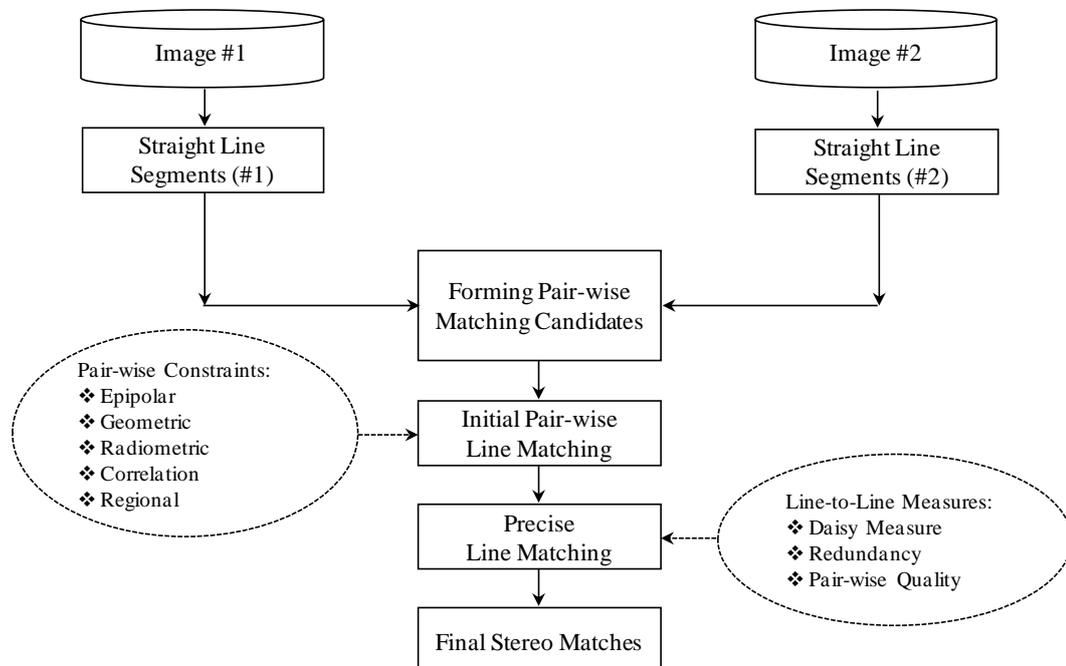


Figure 4.1 Flowchart of the proposed stereo line matching approach

significantly reduced after imposing a weighted pair-wise matching similarity score computed over a total of seven pair-wise constraints (an epipolar, three geometric, a photometric, a correlation and a regional constraint). Thereafter, to select the best line-to-line correspondences, a precise matching step is developed. This step involves newly proposed line-to-line measures (line-based Daisy, Redundancy, Pair-wise Quality) and for each line in the base image, the best corresponding line in the search image is assigned after an iterative final disambiguation process in which the matching inconsistencies are further eliminated using nearest/next distance ratios and a final similarity voting scheme.

During the implementation, for the realization of the image-to-image relations (estimation of the epipolar lines, stereo intersection etc), the well-known photogrammetric techniques are utilized. For this purpose, the precise knowledge of the position of the perspective centers, image orientations and the camera calibration is required. Therefore, for all images, it is assumed that this information is available from a bundle solution (precise exterior, interior and (if available) self-calibration parameters). For this study, we further assumed that the processed stereo images are not significantly different (within  $\pm 5^\circ$ ) in terms of their kappa ( $\kappa$ ) angles. Thus, we do not apply any a priori rotation to the line segments before the matching step.

#### **4.1 Formation of Potential Matching Candidates on the Base Image**

In UHR images, the number of extracted line segments is quite large even for a small part of an urban aerial image. For this reason, in a pair-wise strategy, it is necessary to select reference line pairs in a meaningful manner due to two reasons; (i) to select and search for pairs of lines that have a possible connection in terms of their height values, and (ii) to reduce the time required for the processing of pair-wise matching to a reasonable level. At this point, the former is quite important. The aim is to search for pairs of line segments that have a connection in terms of their height values, and to discard those pair-wise relations

that do not show any reasonable similarity (it is noteworthy to state at this point that, this aim neither requires that the line segments in a pair must have exactly the same height values, nor they must be really intersecting in object space). Since the height values of line segments are not known at this stage, we assess three criteria, (i) proximity, (ii) angle of intersection in image space, (iii) similarity of the radiometric values in the flanking regions, during the selection of the line pairs.

The first measure, proximity ( $T_{\text{prox}}$ ), defines the minimum 2D Euclidean distance ( $d_{ij}$ ) between two lines ( $l_i$  and  $l_j$ ) (Fig. 4.2). It can be defined as a joint minimum of two Euclidean distances: the minimum distance between the endpoints of the line segments in a pair, and the minimum of the orthogonal distances computed from one of the lines to any point on the other line segment. However, it should be noted that in order to accept the orthogonal distance computed, the orthogonal projection must fall exactly on to the line segment. For example in Fig. 4.2, the shortest orthogonal 2D distance between the line segments  $l_1$  and  $l_4$  is selected as the  $d_{14}$  distance. Although the other orthogonal projection line ( $d_{41}$ ) is much smaller than the  $d_{14}$  distance, it is not accepted as a valid distance since the orthogonal projection point does not fall on to the line segment  $l_1$ .

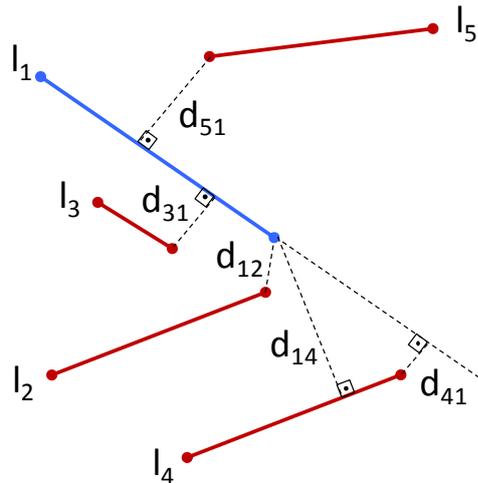


Figure 4.2 The proximity measure for grouping the line pairs on the base image.

The second measure ( $T_{\text{orient}}$ ) is the angle enclosed by line segments  $l_i$  and  $l_j$  (Fig. 4.3). In this part, we only allow formations of line pairs that have a finite intersection point (not parallel) and an angle of intersection value larger than a specific threshold ( $\geq 5^\circ$ ). In Fig. 4.3, the line segments,  $l_1$  and  $l_3$  have approximately similar orientation; therefore, the pair grouping of  $l_1$  and  $l_3$  is not allowed.

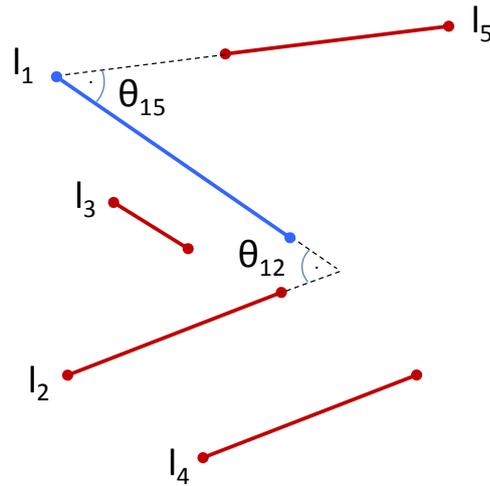


Figure 4.3 The orientation measure for grouping the line pairs on the base image.

The third measure, related to the flanking regions ( $T_{\text{flank}}$ ), is another metric to evaluate the selection of reference line pairs (Fig. 4.4). Apparently, if two line segments in a pair does not confirm any similarity within their flanking regions; those line segments can generally be assumed to be parts of different objects. On the other hand, the line segments that expose large differences within their flanking regions may also have a chance to belong the same object. Thus, the aim at this point is not to eliminate the line pairs that do not show any similarities within their flanking regions, but to compare and learn which side(s) of a pair represents the most similarity. As a result, this information is also held in reserve to be used in the next stage, identifying candidate pair models on the search image.

Apparently, the information of the flanking regions can be computed in several

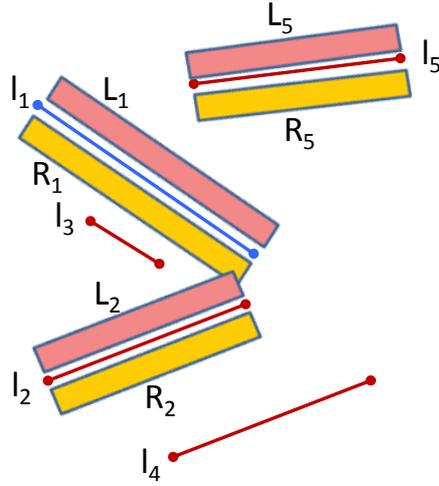


Figure 4.4 The flanking regions measure for grouping the line pairs on the base image.

ways and may involve different color spaces (Henricsson, 1998; Zhang and Baltsavias, 2000); however; here, the multispectral bands are utilized directly by taking the Euclidean norms of the differences of the flanking region information. Thus, a joint minimum of four differences must be computed, i.e., for the line pairs  $l_1$  and  $l_2$  in Fig. 4.4, the flanking region measure can be computed as:

$$T_{flank} = (\|\mathbf{L}_1 - \mathbf{L}_2\|_2; \|\mathbf{L}_1 - \mathbf{R}_2\|_2; \|\mathbf{R}_1 - \mathbf{L}_2\|_2; \|\mathbf{R}_1 - \mathbf{R}_2\|_2)^- \quad (4.1)$$

where,  $\mathbf{L}_i$  and  $\mathbf{R}_i$  represent vectors for mean color values (e.g. RGB) computed for each flanking region in a robust manner (see Chapter 3), and  $(.)^-$  is the joint minimum operator. Note that, if the computed minimum value in Eq. 4.1 is equally shared by a number of difference combinations (e.g. for both  $\|\mathbf{L}_1 - \mathbf{R}_2\|_2$  and  $\|\mathbf{R}_1 - \mathbf{L}_2\|_2$ ), then all those combinations are held in reserve for future testing during the pair-wise matching.

The three measures (proximity, orientation, and flanking regions) used at this stage are not directional, that is the selection of mutual pair relations of ( $l_i$  and  $l_j$ ) or ( $l_j$  and  $l_i$ ) do not have any numerical difference. Therefore, if mutual relations

occur during the processing, we only keep one of the pairs for further processing. This, for sure, reveals a noticeable improvement in terms of speed of the processing; if we have  $n$  lines that have possibility of generating mutual pairs, after the elimination, we have  $n(n-1)/2$  possibilities instead of having  $n(n-1)$ .

#### 4.2 Identification of Candidate Pair Models on the Search Image

Once all the line pairs are selected from the base image, their corresponding matches on the other image are also searched in a pair-wise manner. To fulfill this objective, for each reference pair, all candidate pair models must be collected from the other image. With the knowledge of the image orientations along with the user-specified minimum and maximum height values (or the approximated height information derived from an external DSM data), for a single line, an epipolar quadrilateral region (Fig. 4.5) can be employed to reduce the search space. So far, this regional constraint has been well-known by the photogrammetry and computer vision society and integrated in most of the previous studies dealing with line matching.

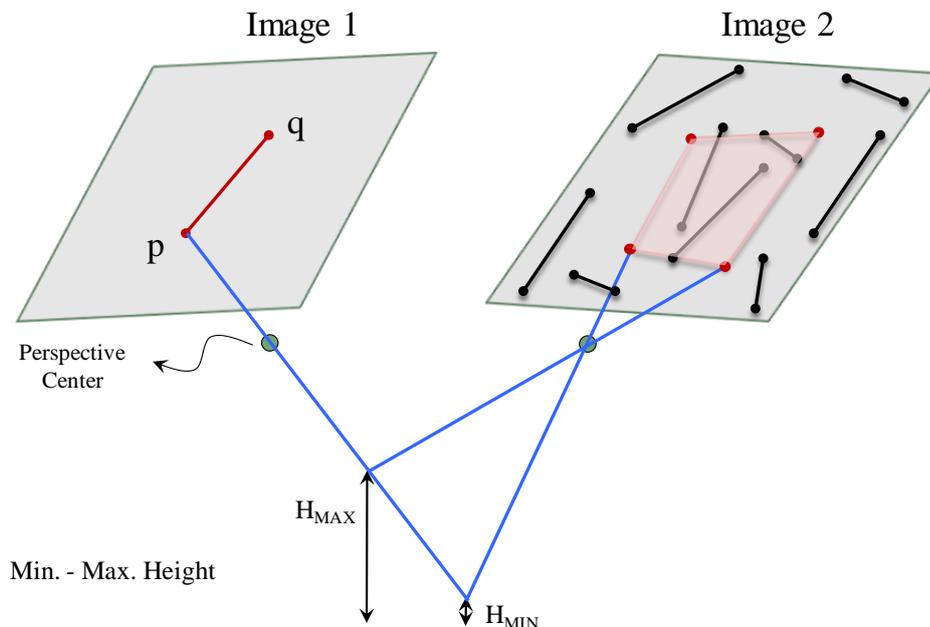


Figure 4.5 The constraint of quadrilateral region to reduce the search space.

On the contrary to the previous studies, in a pair-wise strategy, there are two lines in a pair; thus, all candidates for a pair of lines are collected using two different quadrilateral regions. For example, in Fig. 4.6b, two quadrilateral regions (defined by certain minimum and maximum height) are illustrated for the line pairs,  $l_1$  and  $l_2$  (Fig. 4.6a). However, even for a single line, the number of candidates in each quadrilateral region could be excessive. Here, a constraint is proposed to construct the candidate pair model sets from the individual candidates. As a result, for each line segment, the number of candidates can be considerably reduced. First, the intersection point of the reference pair (Fig. 4.6c) is computed. Since the formation of the reference pairs is restricted in the previous stage with a specific angle ( $T_{\text{orient}}$ , see section 4.1); there is always an intersection point between the line segments that form a reference pair. Thereafter, the epipolar line segment (with the same minimum and maximum heights) of the intersection point on the search image is estimated (Fig. 4.6d). Next, for all candidate pair models, the individual intersection points are computed and the proximity of the points to the epipolar line segment is tested by computing their orthogonal distances. If the distance value is computed to be less than a threshold ( $T_{\text{epi}}$ ), the candidate pair is justified, otherwise deleted. For the threshold  $T_{\text{epi}}$ , rigorous experimental evaluations are performed (see Chapter 6), and it is found that almost all the correct pair intersections are within the range of 5 pixels distance to the epipolar line. Very similar results for the features of junctions are already verified by (Kim and Nevatia, 2004); thus, the relations that have provided orthogonal distances of less than  $T_{\text{epi}}$  are deleted. In Fig. 4.6d, the intersection points of candidate pair models that are computed to be less than  $T_{\text{epi}}$  for the line pair  $l_1$  and  $l_2$  are shown.

Although the epipolar line of intersection constraint is very successful if the lines in a pair actually intersect in object space (or intersect hypothetically), it does not hold for the pairs that are formed by the lines that do not intersect. Thus, the correct pair model (if it exists) on the other image might be missed. For example, in Fig. 4.6a, if we try to force this constraint on a pair formed by the line segment  $l_2$  and any other line that is found on the road surface nearby, the constraint will

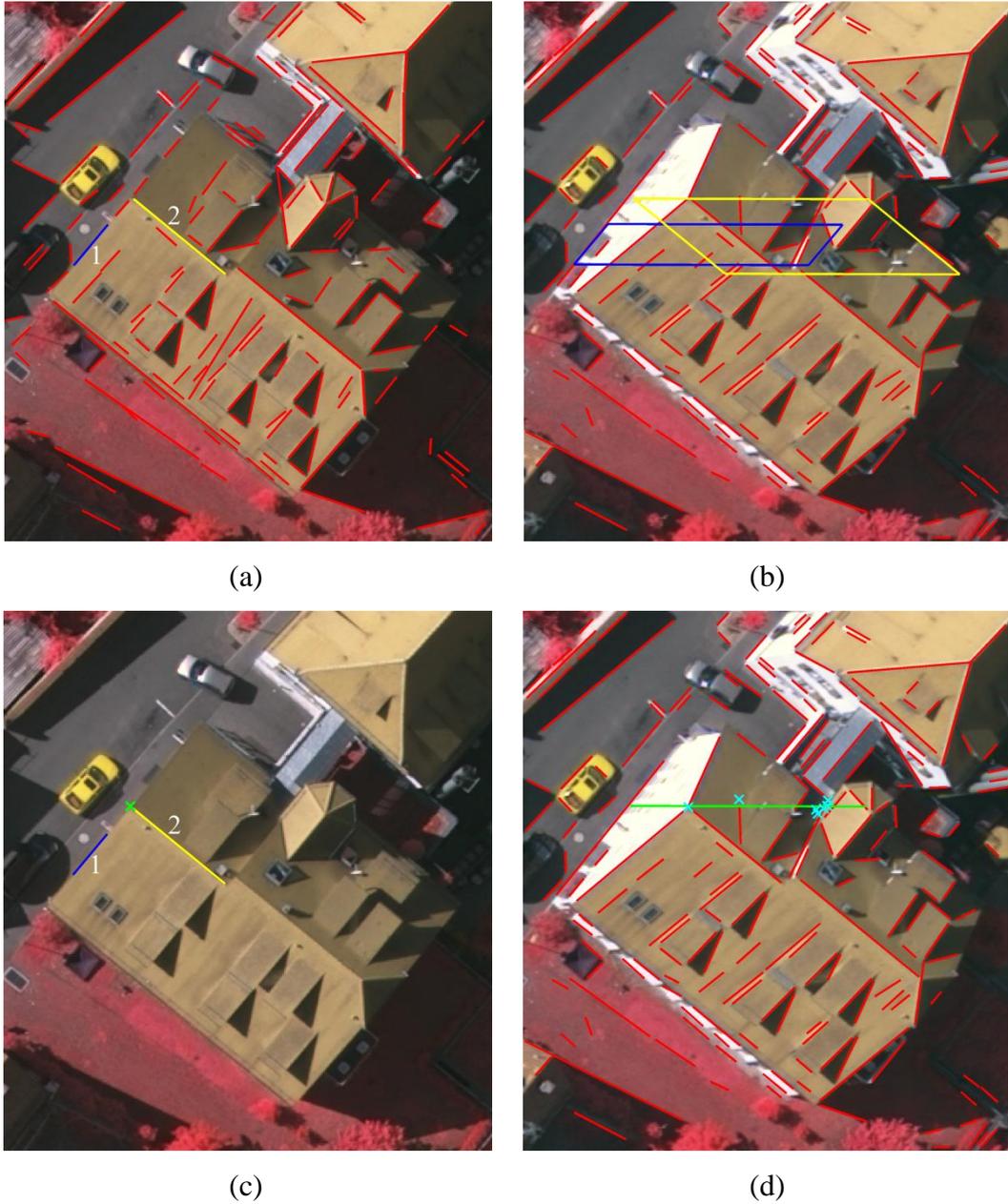


Figure 4.6 (a) Reference line pairs, (b) two quadrilateral regions, (c) the intersection point of  $l_1$  and  $l_2$ , (d) the epipolar line of intersection and the intersections of candidate pair models that are computed to be less than  $T_{\text{epi}}$ .

fail. This is simply due to the reason that, the line segment  $l_2$  has not in height based relation with the segments found on the road surface. However, that kind of pair formations are generally known and can be followed from the pair formation step with the information obtained through the flanking regions of the line

segments (see section 4.1). One different aspect of this constraint is that it also automatically eliminates the reference pairs in which two line segments in one view correspond to a single segment in the other view (lack of a unique intersection point). For instance, in Fig. 4.6a, assume that the line segment  $l_1$  is extracted much longer in a case which discriminates the entire building roof completely from the road segment. In this case, a single segment,  $l_1$ , will simply correspond to two different line segments on the second view. To solve these cases, the repetitive nature of the pair formation is used considering the fact that a single line is allowed to have a part in different pair models. Thus, a single line has possibility to be matched with its correct correspondences in different pair models; therefore in such a case,  $l_1$  has possibility to be matched with its two separate correspondences in different combinations of reference pair models.

At the end of this step, for each reference pair model, all possible candidate pair models are collected. Thus, the pair-wise relations between the entire reference and candidate pairs are established and ready for the subsequent matching stage.

### **4.3 Initial Pair-wise Matching**

In this part, new pair-wise constraints are developed and utilized for the initial stereo pair-wise line matching scheme. In particular, a total of seven pair-wise constraints (an epipolar, three geometric, one photometric, a correlation and a spatiogram constraint) from different domains (geometric, radiometric, and regional) are developed. These constraints are integrated in a weighted pair-wise matching similarity function to select the best matching candidate pair for each reference line pair generated from the base image.

#### **4.3.1 Geometric Constraints**

In order to describe the geometrical relations between the line segments in a pair, we employ three different measures (Park et. al., 2000). The first measure is the

angle which two line segments  $l_1$  and  $l_2$  form (Fig. 4.7a), the second measure is the angle from the midpoint of  $l_1$  to that of  $l_2$ , which is measured from the first line to the second line (Fig. 4.7a), and the third measure is the ratio of the sum of the line lengths of segments to the average distance between the endpoints of the line segments (Fig. 4.7b). Thus, the geometrical relations utilized are defined as (Park et. al., 2000):

$$\begin{aligned}
 r_1(l_1, l_2) &= \theta_{11} \\
 r_2(l_1, l_2) &= \theta_{21} \\
 r_3(l_1, l_2) &= \frac{1}{d}(\overline{AB} + \overline{CD})
 \end{aligned}
 \tag{4.2}$$

where  $d = (\overline{AC} + \overline{AD} + \overline{BC} + \overline{BD}) / 4$ .

In the line matching literature, these types of geometric constraints were already known and utilized (eg., Park et. al., 2000, Zhang and Baltasvias, 2000). However, in most of the cases, two geometrical relations ( $r_2$  and  $r_3$ ) are inappropriate for most of the aerial images, since the line segments found in different views may have different lengths and midpoints due to several reasons such as occlusion, image noise etc. In addition, the perspective distortion

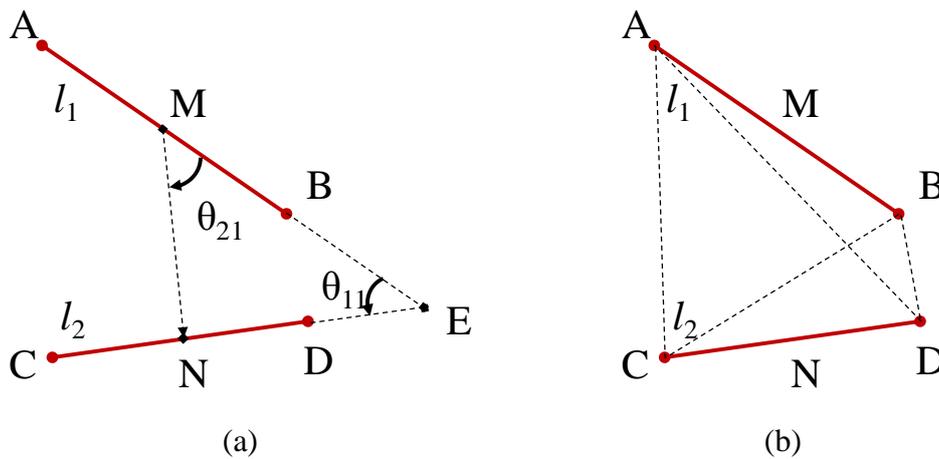


Figure 4.7 The geometrical measures utilized.

combined with relief of terrain and/or of individual objects also plays an important role at this point and in summary, the measures completely become inconsistent from one view to another.

Assume that the lines  $c_1$  and  $c_2$  in Fig. 4.8b form one of the candidate pair models of the lines  $l_1$  and  $l_2$  in Fig. 4.8a. If we compare the lengths of the lines in each pair, only the length of the line  $c_2$  is significantly different; however, even in this case, two geometrical measures computed are different from each other. For this reason, a normalization scheme is developed to deal with the problems of the geometrical reliability of the line segments extracted from different views. It relies on the epipolar geometry and the idea of finding the common overlaps of lines in different views. For the endpoints of each line, the corresponding epipolar lines on the other view are estimated. Thus, a point to point correspondence (Schmid and

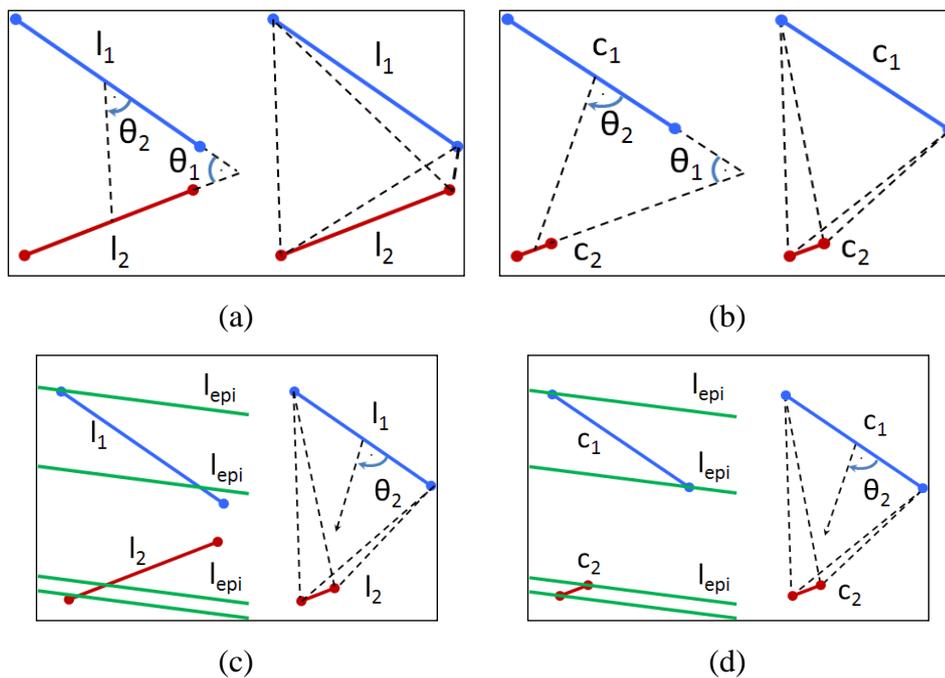


Figure 4.8 (a) A reference model, (b) a candidate pair model and its geometrical measures, (c) and (d) normalization with epipolar lines and the normalized measures.

Zisserman, 1997) is performed on each line to provide a final single overlapping line for each line in a pair (Fig. 4.8c- and 4.8d-left). This normalization scheme for each reference pair and its candidate pair model is also applied before the computation of the second and third geometrical measures. Thus, the measures turn out to be more reliable (Fig. 4.8c- and 4.8d-right) when compared to their non-normalized counterparts.

Once the geometrical relations are computed, we define the geometrical constraints between a reference pair and a candidate pair by taking the absolute differences:

$$\xi_i = |r_i(l_1, l_2) - r_i(c_1, c_2)| \quad i \in \{1, 2, 3\} \quad (4.3)$$

### 4.3.2 Radiometric Constraint

In this study, a flanking region constraint that searches for the intra-pair similarity between the reference pairs and the candidate pairs are developed (Fig. 4.9). The constraint takes into account the similarity of the side(s) of the reference pair model previously found (Eq. 4.1) and searches whether a similar relationship of the flanking information of the same sides for the candidate pair models exist or not. To allow such a constraint, the illumination of the images is assumed to be similar (the case in a single strip acquisition) and the reflections are assumed to comply with the lambertian theory.

Assuming that the most similar flanking regions in a reference pair is computed as  $\|\mathbf{L}_1 - \mathbf{R}_2\|_2$ , the Eq. 4.1 becomes:

$$\begin{aligned} T_{flank}^R &= (\|\mathbf{L}_1 - \mathbf{L}_2\|_2; \|\mathbf{L}_1 - \mathbf{R}_2\|_2; \|\mathbf{R}_1 - \mathbf{L}_2\|_2; \|\mathbf{R}_1 - \mathbf{R}_2\|_2)^- \\ &= \|\mathbf{L}_{R1} - \mathbf{R}_{R2}\|_2 \end{aligned} \quad (4.4)$$

Subsequently, a similar relationship of the flanking information of the same sides

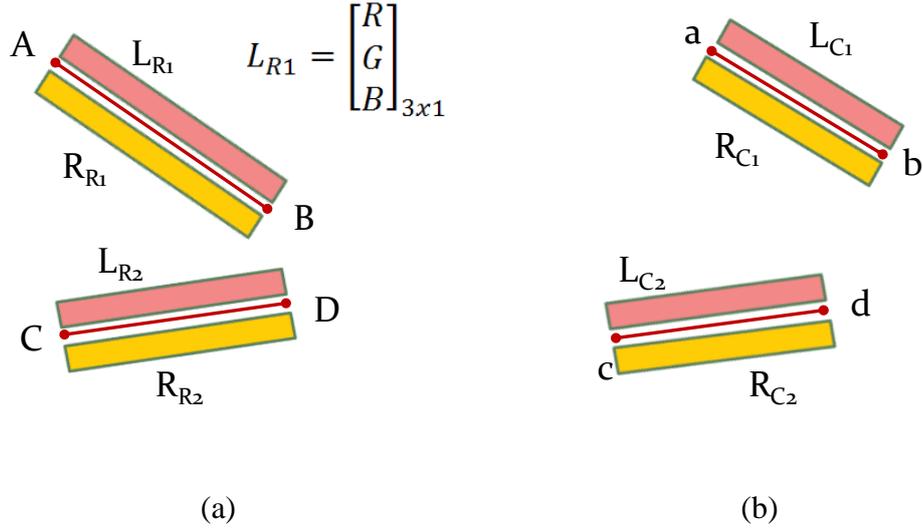


Figure 4.9 (a) A reference pair and (b) a candidate pair with flanking regions.

for the candidate pair model must hold:

$$T_{flank}^C = \|\mathbf{L}_{C1} - \mathbf{R}_{C2}\|_2 \quad (4.5)$$

Thus, the intra-pair similarity constraint between the two flanking regions can be defined as:

$$\xi_4 = |T_{flank}^R - T_{flank}^C| \quad (4.6)$$

### 4.3.3 Correlation Constraint

In order to further assess the similarities between the reference and candidate pairs, a pair-wise correlation constraint forced on a hypothesized 3D triangular plane is developed. The correlation constraint performs on a 3D plane fitted to the line pairs and their intersection point based on the assumptions that (i) they are the correct match, and (ii) they belong to a single plane. A correlation measure bounded for all the area marked by the 3D lines and the intersection point is not appropriate, since there may be different planes on a building roof (chimneys, dormers etc). Thus, we apply the correlation measure to the immediate vicinity of

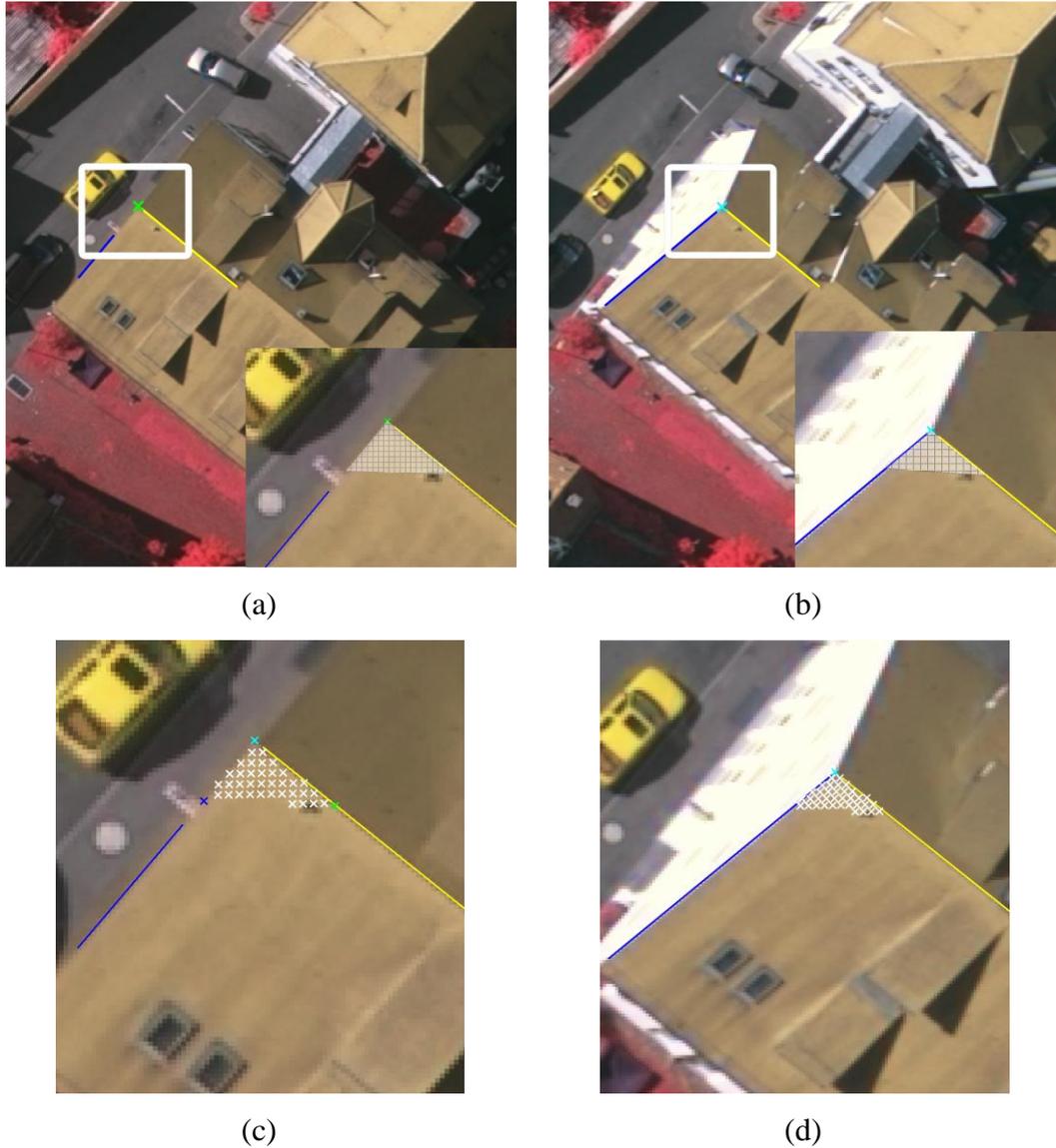


Figure 4.10 (a, b) The back projected plane estimated from the line pairs and the intersection point, (c, d) points that are utilized for the one-to-one correspondence based on the height values obtained from the estimated 3D plane.

the point of intersection and the corresponding plane, which can also be defined as a 3D triangular planar patch (Fig. 4.10). We fixed the side lengths of the triangle which are exactly on the same direction of the lines by a single distance parameter  $d = 2$  m. Fig. 4.10a and Fig.4.10b illustrate the extents of the back projected plane that is estimated through the line pairs given in the figure. Next, the 3D surface of the plane is then utilized to perform point-to-point correspondence between the

two images (Fig. 4.10c and Fig.4.10d). Thus, a correlation value can be computed between the two images by collecting the pixel values (nearest-neighbor re-sampling). At this point, the original image is converted to the Lab color space (CIE, 1976) and the luminance band of the Lab space is utilized during the computation of the correlation values. Thus, if we have  $m$  distinct points on the surface of the 3D plane, two vectors that consist of  $m$  values can be generated for the base and the search image:

$$\mathbf{R} = [L_1^R \quad L_2^R \quad \dots \quad L_m^R]^T \quad \mathbf{C} = [L_1^C \quad L_2^C \quad \dots \quad L_m^C]^T \quad (4.7)$$

where  $L$  represents the Luminance values collected,  $\mathbf{R}$  and  $\mathbf{C}$  denotes reference and candidate pairs, respectively. Thus, the correlation constraint between the two vectors can be computed as:

$$\xi_5 = \rho_{R,C} = \frac{E[(R-\mu_R)(C-\mu_C)]}{\sigma_R\sigma_C} \quad (4.8)$$

In Eq. 4.8,  $\rho$  defines the correlation coefficient,  $E$  is the expected value operator,  $\mu$  and  $\sigma$  denote the mean and standard deviations, respectively. However, it should be noted that there may be several cases that may violate the plane formation and the correlation value computed: (i) the intersection point of the lines that are exactly on the same plane may appear to be on a different plane(s) than their own plane (Fig. 4.11a – pair<sub>AB</sub>) (ii), the lines that really intersect in object space may not correspond to a physical object plane (Fig. 4.11a – pair<sub>CD</sub>), and (iii) the planes formed by the line pairs may be hidden or occluded in one view (Fig. 4.11b – pair<sub>EF</sub>). It is straightforward to track the last violation; we compute the angle difference of the plane with its projected plane (to a flat terrain), and only apply the correlation measure if the computed plane angle is narrower than a specific angle threshold ( $\leq 75^\circ$ ). However, the other two violations cannot be handled in a similar manner, since the hypothesized 3D planes are not correct. For that reason, it is important to note that some of the correct pair relations may also produce low

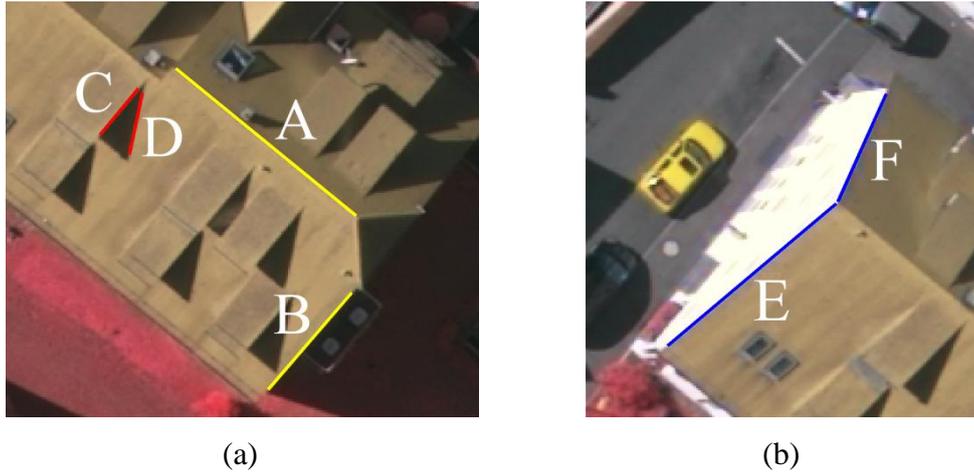
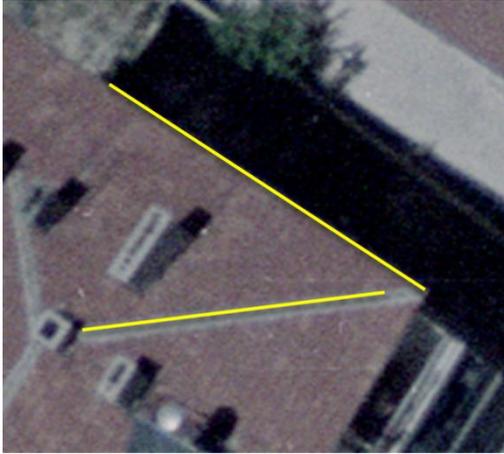


Figure 4.11 (a, b) Several cases that may violate the plane constraint.

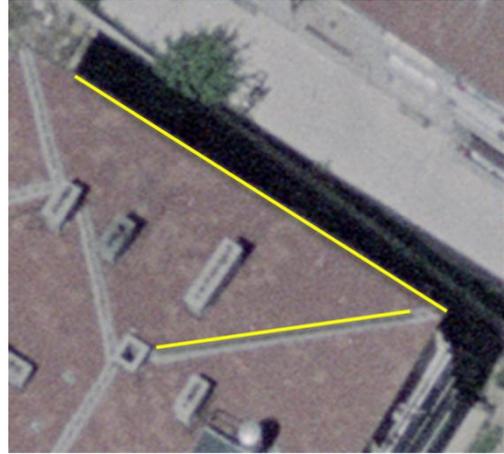
correlation scores. Therefore, the threshold value utilized for the constraint is carefully analyzed and determined after performing a detailed investigation (see details in Chapter 6).

#### 4.3.4 Regional Constraint

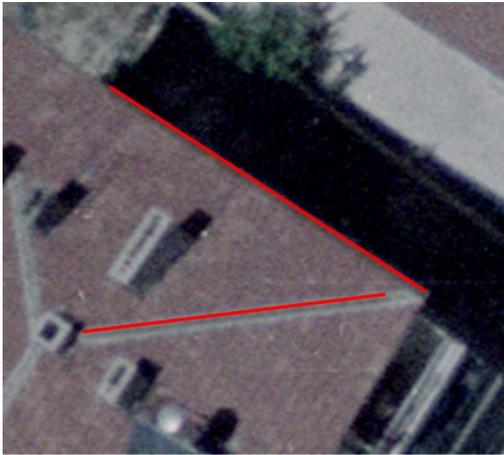
The regional similarities dominated by the reference line pairs and the candidate pair models are also evaluated. To do that, first, the 2D regions that are consistently described by the line pairs are selected. Fig. 4.12 illustrates the selection procedure. Initially, for each reference and the candidate line pair, point-to-point correspondence is applied to obtain consistent line end-point locations (Fig. 4.12 c, d). Next, the regions dominated by those line segments are selected (Fig. 4.12 e, f). However, it is simply impossible to compare the regions directly, since the perspective distortion and the features belong to many different planes on the roofs (e.g., chimneys, dormers etc.) may simply alter the positions of the pixels to some extent. On the other hand, it is not logical to compare the regions with a simple histogram measure, since many parts of the images may contain similar radiometric information; very different regions generally produce similar histograms. Therefore, we utilize the spatiogram measure (Ó Conaire et. al., 2007) to evaluate the regional similarity between the regions. A very important aspect of



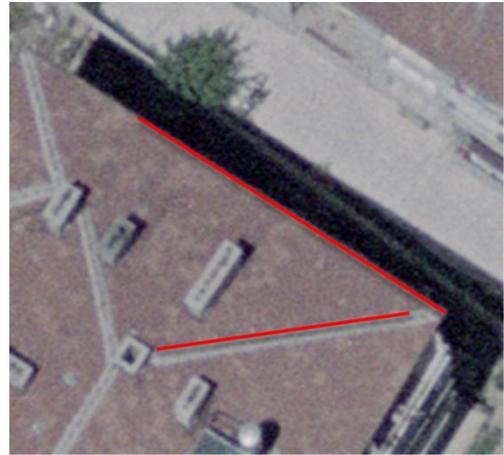
(a)



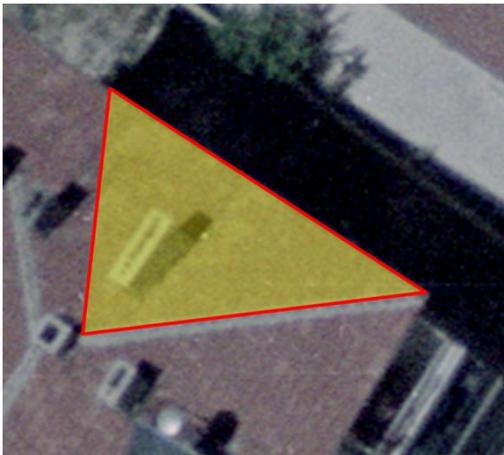
(b)



(c)



(d)



(e)



(f)

Figure 4.12 (a, b) Corresponding line pairs, (c, d) same line pairs after point-to-point correspondence, (e, f) regions dominated by the line pairs.

the spatiogram measure is that it has a unique capability to combine the distribution of the radiometric information along with the spatial information. Thus, the positional differences occur between the line pairs are handled (somewhat alleviated) while providing the histogram information. The details of the derivation of the spatiogram measure are explained in (Ó Conaire et. al., 2007); in this part, only the final similarity measure is provided:

$$\xi_6 = \sum_{b=1}^B \sqrt{n_b n'_b} \left[ 8\pi |\Sigma_b \Sigma'_b|^{1/4} N(\mu_b; \mu'_b, 2(\Sigma_b + \Sigma'_b)) \right], \quad (4.9)$$

where,  $B$  is the number of histogram bins utilized,  $N(x; \mu, \Sigma)$  operator represents a normalized Gaussian evaluated at  $x$ ,  $\{n_b, \mu_b, \text{ and } \Sigma_b\}$  and  $\{n'_b, \mu'_b, \text{ and } \Sigma'_b\}$  are the two spatiograms extracted from the base and the search images with the specific histogram parameters of bin count, spatial mean and spatial covariance, respectively.

#### 4.3.5 Epipolar Constraint

In section 4.1, during the formation of candidate pairs, to limit and reduce the number of candidate pair models, an epipolar line based approach is proposed. The approach initially generates an intersection point between the line segments that form a reference pair. Thereafter, the epipolar line segment (with certain minimum and maximum heights) of the intersection point on the search image is estimated. Next, the proximity of each intersection point to the epipolar line segment is tested by computing the orthogonal distances ( $d_{\perp}$ ) between the point and the epipolar line segment (Fig. 4.13). As a result, this orthogonal distance can also be utilized as a new measure, since we search for pairs of line segments that confirm a relation in terms of their height values (more strictly, intersecting line segments in object space). Thus, if the computed distance for a candidate line pair is relatively shorter than the distances of the other candidate line pairs, it's clear that the pair with the shortest distance is more likely to be the correct matching pair. Thus, if  $(x_1, y_1; x_2, y_2)$  represents two arbitrary points on the epipolar line

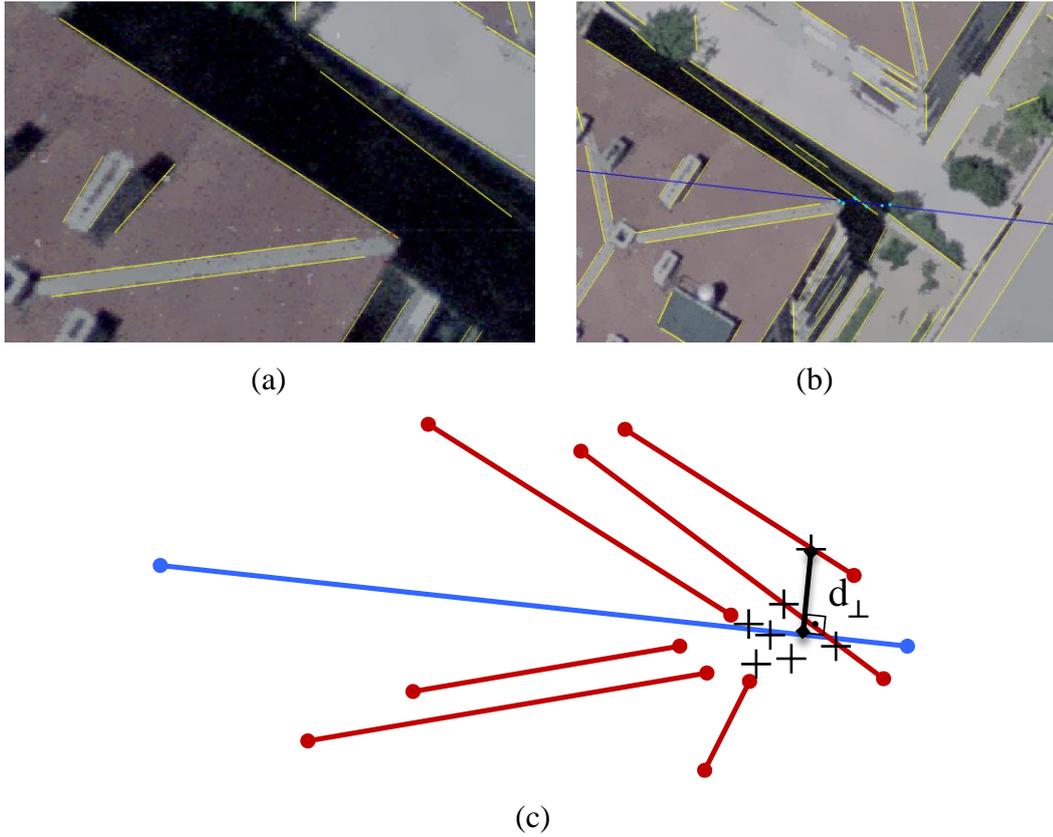


Figure 4.13 (a) A reference pair selected from the base image and the intersection point (in red color), (b) the estimated epipolar line of the intersection point on the search image (in dark blue color) and the possible intersection points (in light blue color) of candidate pairs around the epipolar line, and (c) the orthogonal distance between the intersection points and the epipolar line.

segment and  $(x_i, y_i)$  is the image coordinates of the intersection point; a similarity measure can be defined as:

$$\xi_7 = d_{\perp} = \sqrt{(x_i - x_1 - u \times (x_2 - x_1))^2 + (y_i - y_1 - u \times (y_2 - y_1))^2}$$

where (4.10)

$$u = \frac{(x_i - x_1) \times (x_2 - x_1) + (y_i - y_1) \times (y_2 - y_1)}{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

#### 4.3.6 Pair-wise Matching

The final pair matches are assigned after a weighted pair-wise matching similarity score which is computed over a total of seven measures ( $\xi_1 - \xi_7$ ); an epipolar, three geometric, one photometric, a correlation and a regional measure. However, it is clear from those measures that, except for the correlation and regional scores ( $\xi_5$  and  $\xi_6$ ), the scores of the other five measures define the dissimilarities between the two pair relations. Therefore, prior to the computation of overall pair-wise similarities, the scores of all measures are computed as dissimilarities by normalizing each measure (by means of min. and max. values) from 0 to 1, and the overall similarity result ( $\theta_s$ ) between the two pair relations is computed as:

$$\theta_s = \left( \frac{1}{1 + \sum_{i=1}^7 \xi_i^n w_i} \right)^2 \quad (4.11)$$

In Eq. 4.11,  $\xi_i^n$  is the normalized dissimilarity value of each measure, and  $w_i$  represents the weights associated with each measure. During the computation of the overall similarities, the weights utilized for each measure are selected as equal ( $w_i = 1/7, \forall i$ ) and the computed similarities are squared (Eq. 4.11) in order to give more weight to high similarity values, and to be more selective.

#### 4.4 Precise Line-to-Line Matching

Once the initial pair-wise matches are established, the line-to-line relations between the stereo images can be inferred. However, the results of the pair-wise matching rarely provide one-to-one line correspondences between the images. Based on our experiences, after the pair-wise matching, the ambiguities mostly occur for the lines that are adjacently located within a very short perpendicular distance. Typical examples can be seen in Fig. 4.14c, d. This is mainly due to two

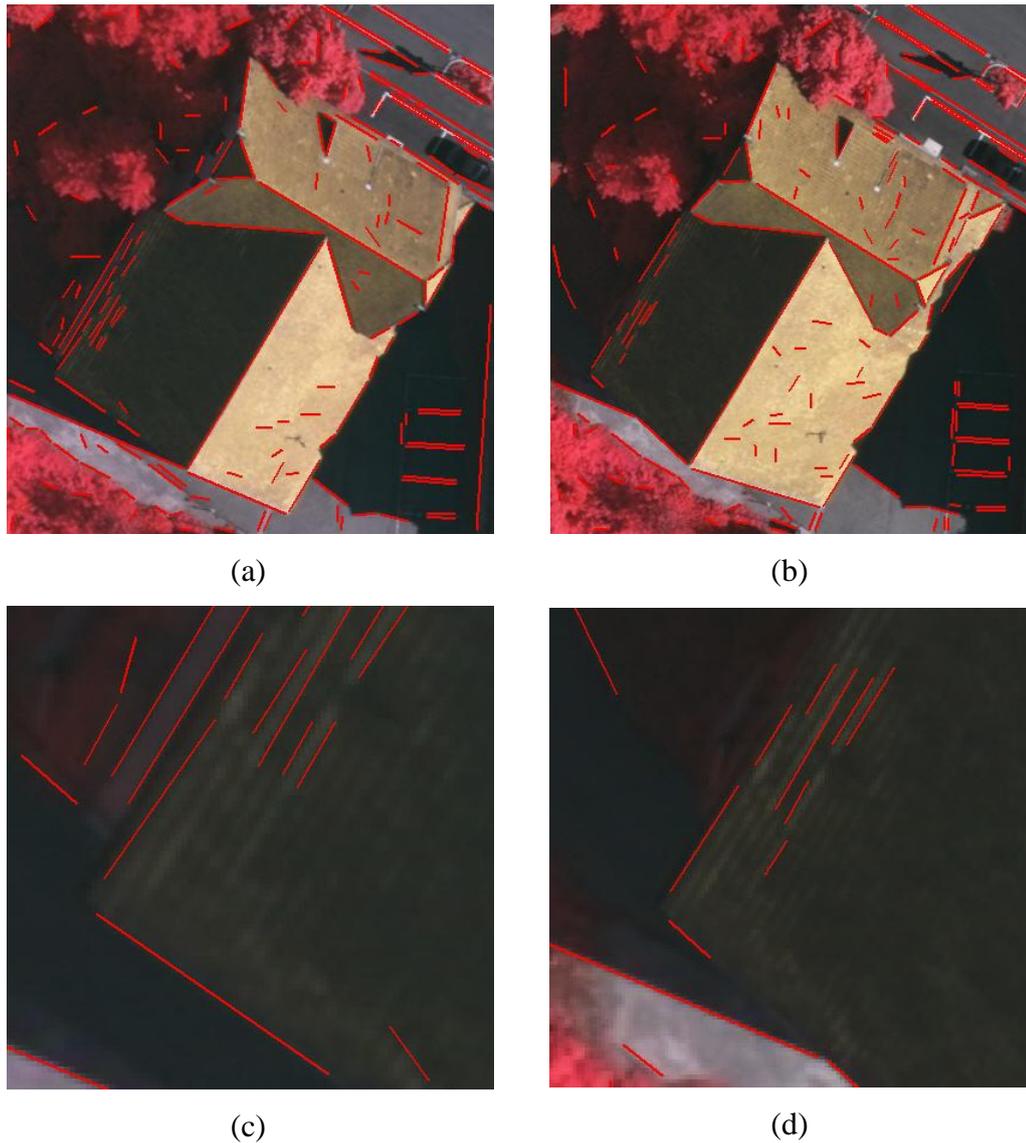


Figure 4.14 The matching ambiguities after pair-wise matching. (a, b) The stereo images and the extracted lines, (c, d) potential ambiguities after the pair-wise matching.

explicit reasons; (i) the lines that are very close to each other that belong to the same object (building, road etc.) reveal similar pair-wise characteristics and (ii) since a relaxed threshold ( $T_{\text{epi}}$ ) is applied for the epipolar intersection during the candidate pair formation, very close lines are mostly susceptible to satisfy that threshold. Therefore, in this thesis, a great care has been devoted to the precise line-to-line matching stage and a new iterative disambiguation algorithm is

developed. For this purpose, we combined three novel measures during the selection of the best line correspondences. (i) The first measure relies on the gradient orientation information in the local neighborhood of lines which is computed using a recently proposed dense matching measure, Daisy (Tola et. al., 2010). Since the original Daisy measure is point based, in this thesis, the measure is extended and adapted to fulfill the requirements of the linear features and their local neighborhood. (ii) The second measure, the Redundancy, is computed from the entire pair-wise matches based on the fact that a single line is allowed to have a part in different pair combinations. Thus, after the pair-wise matching, there is a quite large number of matching redundancy available for most of the line correspondences. By this way, the redundancy measure gives a possibility to understand and integrate exclusive local matching information for line segments during the disambiguation process. (iii) The third measure is computed from the results of each individual pair-wise matching. Since we assigned the best pair using a pair-wise matching similarity score, this information can also be utilized during the precise matching stage, since the quality of the pair matches are inherently determined by the quality of the line correspondences in each pair. Those three measures are integrated for the final disambiguation process in a newly developed iterative way, in which the matching inconsistencies are eliminated using nearest/next ratios and a final similarity voting scheme.

#### **4.4.1 Line-based Daisy Measure**

In recent years, the gradient orientation histograms has proven to be robust to distortions (up to a level) and found to be successful in terms of point matching when compared to the classical pixel-based measures such as cross-correlation and pixel differencing. Some good examples can be found in (Lowe, 2004; Mikolajczyk and Schmid, 2005; Bay et. al., 2006). More recently, Tola et. al. (2010) proposed a dense matcher, Daisy, which has also proven to be much more efficient during the computation of the gradient orientation histograms. In the line matching context, up to our knowledge, the only study that takes into account the

gradient orientation around the line local neighborhood was proposed by Wang et. al. (2009). However, for aerial images, the final matching results that are only based on the information obtained from the line local neighborhoods could be ambiguous. Because all those measures are almost non-discriminative on their own for the aerial image case and suffer from the same problem, repetitive patterns, where the information extracted from local neighborhoods of very different lines has similar information (Fig. 4.15). However, the information within those neighborhoods may reveal some hints and may provide opportunities to indicate and eliminate the indisputably wrong matches.

In this study, we selected the Daisy as a fundamental local neighborhood measure for the post-processing due to two explicit reasons: (i) great efficiency and speed during the computation of the gradient orientation histograms, and (ii) its circular, symmetric shape and isotropic kernel structure turns out in a small overhead during the computation of the measure for different line orientations. Here, first, we only briefly review the original point-based Daisy measure and refer the reader to the reference (Tola et. al., 2010) for further details. Thereafter, we will introduce new adaptations for Daisy and present how efficiently the measure

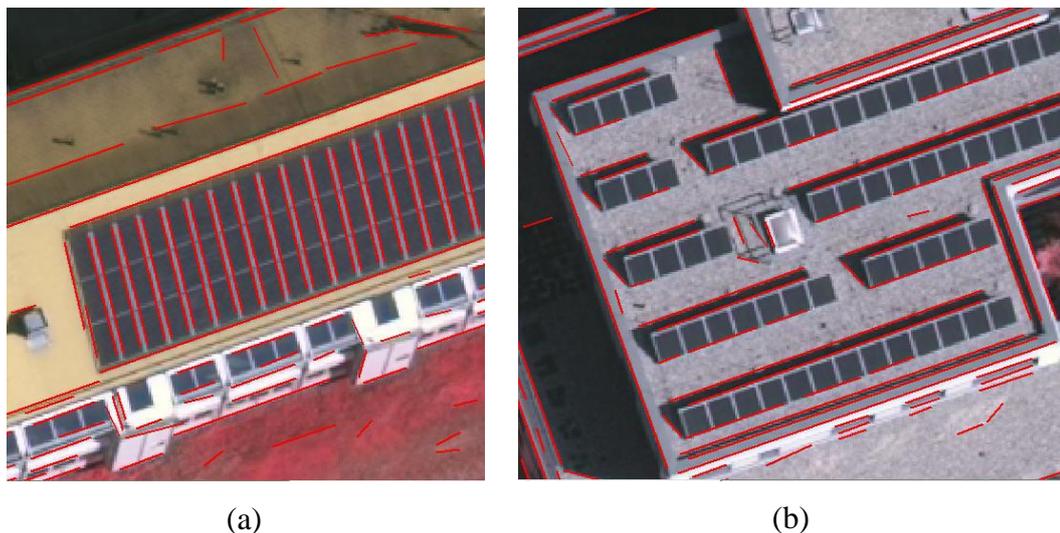


Figure 4.15 Two examples of line segments commonly observed in repetitive patterns and their very similar local neighborhoods.

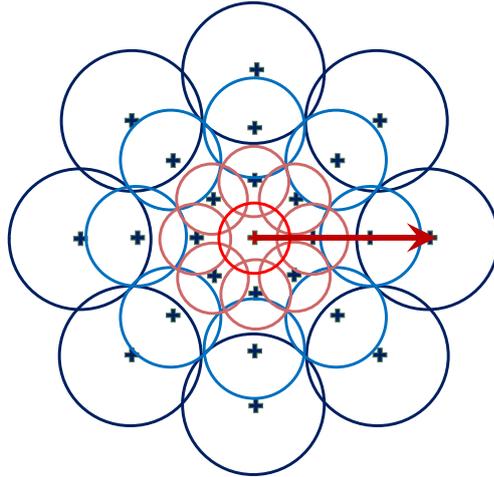


Figure 4.16 The original Daisy descriptor (Tola et. al., 2010).

could be utilized for capturing the local line neighborhoods.

The Daisy descriptor is given in Fig. 4.16. In the descriptor, each circle represents a region where the radius is proportional to the standard deviations of the Gaussian kernels and the “+” sign represents the pixel locations of the convolved orientation map centers where the descriptor is computed. Daisy is controlled by a total of four parameters; where  $R$  is the distance from the center pixel to the outer most grid point,  $Q$  is the number of convolved orientation levels,  $T$  is the number of histograms at a single layer, and  $H$  is the number of bins in the histogram. For a given input image, first, depending on the number of bins  $H$ , orientation maps are computed. Each orientation map is then incrementally convolved with Gaussian kernels of different sigma values to obtain convolved orientation maps. At each pixel location illustrated in Fig. 4.16, a vector made of values from the convolved orientation maps is computed. Let  $\mathbf{h}_{\Sigma}(u,v)$  represent the vector made of the gradient orientation values at location  $(u,v)$  in the orientation maps after convolution by a Gaussian kernel of standard deviation  $\Sigma$ , and let  $Q$  represents the number of different circular layers, then the Daisy descriptor  $D(u_0,v_0)$  for location  $(u_0,v_0)$  is defined as (Tola et. al., 2010):

$$\begin{aligned}
D(u_0, v_0) = & \left[ \tilde{\mathbf{h}}_{\Sigma_1}^T(u_0, v_0), \right. \\
& \tilde{\mathbf{h}}_{\Sigma_1}^T(\mathbf{l}_1(u_0, v_0, R_1)), \dots, \tilde{\mathbf{h}}_{\Sigma_1}^T(\mathbf{l}_T(u_0, v_0, R_1)), \\
& \tilde{\mathbf{h}}_{\Sigma_2}^T(\mathbf{l}_1(u_0, v_0, R_2)), \dots, \tilde{\mathbf{h}}_{\Sigma_2}^T(\mathbf{l}_T(u_0, v_0, R_2)), \\
& \dots \\
& \left. \tilde{\mathbf{h}}_{\Sigma_Q}^T(\mathbf{l}_1(u_0, v_0, R_Q)), \dots, \tilde{\mathbf{h}}_{\Sigma_Q}^T(\mathbf{l}_T(u_0, v_0, R_Q)) \right]^T, \tag{4.12}
\end{aligned}$$

where  $\mathbf{l}_j(u, v, R)$  is the location with distance  $R$  from  $(u, v)$  in the direction given by  $j$  when the directions are quantized into the  $T$  values.

Since the Daisy descriptor is point-based (it belongs to the center grid point), in this study, the measure is extended and adapted to fulfill the requirements of the line features and their local neighborhood. First, we centralize the center grid point of the descriptor to the center of the overlapping parts of the line segments which are defined by point to point correspondence (Fig. 4.17 a, b). Next, to achieve rotation invariance over gradient vectors, we rotate the Daisy grid and align the direction vector of the descriptor with the orientation of each line (Fig. 4.17 c, d). Since the amount of rotation must be adjusted for all lines based on their angle values in image space, during this procedure, we fully utilize one of the main advantages of the Daisy descriptor in which we only circularly shift the final orientation histograms to compute the descriptor. To achieve invariance to perspective distortion exactly on the line segments, for each line, we utilize adaptive  $R$  values for the Daisy grid (distance from the center pixel to the outer most grid point). The original Daisy measure has a specific constant  $R$  value; however, adaptive  $R$  values for line segments could be computed with the knowledge of the overlapping parts after imposing point to point correspondence. Since we apply this correspondence during the initial pair-wise matching, it does not bring any further overhead during the computation of the measure. In addition, it is apparent that we don't have any knowledge about the surfaces attached to the lines in their neighborhoods; thus, we further utilize the adaptively computed  $R$  values for entire Daisy grid points (Fig. 4.17 e, f). After these adaptations, for the computation of the similarities, we divide the Daisy grid points into two separate

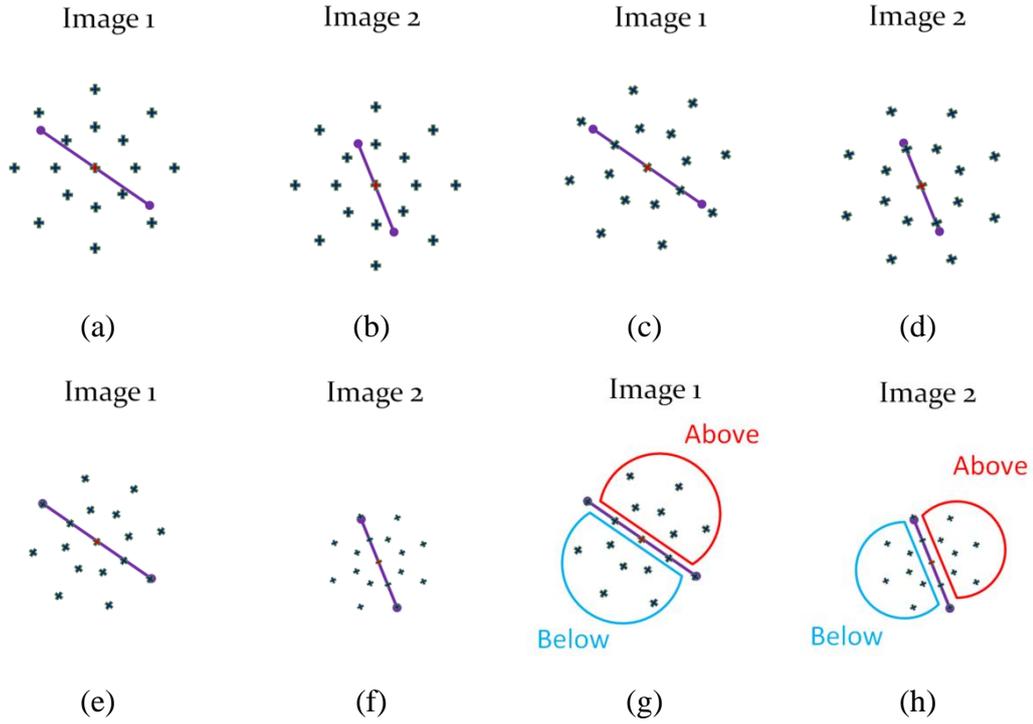


Figure 4.17 The adaptation of the Daisy descriptor. (a, b) The line segments with point-to-point correspondence and the located Daisy grids to the center of each line segment, (c, d) rotated grids based on the orientation of the line segments, and (e, f) final Daisy grids with adaptive scaling, (g, h) divided Daisy grid points.

classes and produce two constant grid binary masks  $\{M_m(x)\}$  for each line; the grid points that are located (i) above the line, and (ii) below the line (Fig. 4.17 g, h). Thus, we perform the similarity computations independently for each grid class. Moreover, we also mask out the vectors,  $\mathbf{h}_{\Sigma}(u, v)$ , from the descriptor matrix  $D(u_0, v_0)$  whose grid locations are exactly on the line. This is due to the reason that if one of the sides of the lines is occluded, then the histograms computed for the points that are exactly on the lines have no reason to resemble each other. Therefore, we exclude those pixel locations and their histograms from the Daisy measure. For the computation of the dissimilarities between two Daisy descriptors, Tola et. al. (2010) proposed a Euclidean difference metric;

$$D = \frac{1}{\sum_{q=1}^S M^{[q]}} \sum_{k=1}^S M^{[k]} \left\| D_i^{[k]}(x) - D_j^{[k]}(x) \right\|_2, \quad (4.13)$$

where  $S$  is the number of grid points,  $M^{[k]}$  is the  $k$ th element of the binary mask  $M$ , and  $D_i^{[k]}$  is the  $k$ th histogram  $\mathbf{h}$  in  $D(x)$  computed from image  $i$ . However, we observed that, although the metric is successful in most of the cases, it completely ignores the cross-correlation between the two descriptors,  $D_i$  and  $D_j$ . Thus, we define a modified-similarity ( $M_S$ ) metric that can be jointly utilized with the cross-correlation:

$$M_S = \frac{1}{1 + \left( \sum_{k=1}^S M^{[k]} \|D_i^{[k]}(x) - D_j^{[k]}(x)\|_2 \right)^2} \quad (4.14)$$

First, the normalization coefficient in Eq. 4.13 is not necessary any longer since our binary masks have constant number of points for each side of the lines. After the modification, the similarity metric produces values between 0 and 1, and in order to be more discriminative, we take the square of the total dissimilarity, thus, we further penalize the higher dissimilarities values ( $D > 1$ ) and give more weight on the lower ones. We also define the cross-correlation-similarity ( $C_S$ ) between two descriptors as:

$$C_S = \left( \frac{\sum_{k=1}^S (M^{[k]} D_i^{[k]}(x) - \mu_{(MD_i)}) (M^{[k]} D_j^{[k]}(x) - \mu_{(MD_j)})}{s(MD_i) s(MD_j)} \right)^2, \quad (4.15)$$

where  $\mu(\cdot)$  and  $s(\cdot)$  operators denote the mean and standard deviations, respectively. Note that, in Eq. 4.15, similar to Eq. 4.13, the correlation is also squared in order to give more weight to high similarity values, and to be even more discriminative. Finally, since the similarities in Eq. 4.13 and 4.14 are computed independently for both sides of lines (for the above and below grid points); we propose the final line-based Daisy similarity ( $Sim_D$ ) for line matching as:

$$Sim_D = \left( (M_S^{Above}, M_S^{Below})^+, (C_S^{Above}, C_S^{Below})^+ \right)^-, \quad (4.16)$$

where  $(.)^+$  and  $(.)^-$  denote  $\max(a, b)$  and  $\min(a, b)$ , respectively. The final  $(.)^-$  operator in Eq. 4.16 ensures that the final Daisy similarity metric ( $Sim_D$ ) should be high for both of the similarity metrics,  $M_S$  and  $C_S$ .

#### 4.4.2 Redundancy Measure

In this thesis, a new measure, Redundancy, is proposed to solve the matching ambiguities. The Redundancy is computed from the entire pair-wise matches based on the fact that a single line is allowed to have a part in different pair combinations. Thus, after the pair-wise matching, for each line segment, there is a number of matching redundancy that could be efficiently utilized for the disambiguation. The idea is illustrated in Fig. 4.18. Assume that Fig. 4.18 shows the line segments extracted from two stereo images, and Table 4.1a gives the results of the pair-wise matches for those line segments that only had a pair-wise relation with segment #1. In Table 4.1a, the left column represents the search pairs generated from the left image, and the right column represents the best pair matches assigned after the pair-wise matching. If we look at the results of the pair-wise matches in detail, the segment #1 had a total of five pair-wise relations with other line segments within the pre-defined proximity (see part 4.1) in image space. One-to-one line matches inferred from the pair-wise relations are given in Table 4.1b. Based on the uniqueness constraint, a single line segment from the left image has, at most, one corresponding line segment in the right image (Suveg and Vosselman, 2004). However, during the extraction of line segments, a single segment is often fragmented into several shorter segments, such as in the case for the segments #2 and #3. In this case, the uniqueness constraint must be handled carefully by taking into account the collinearity of the fragmented ones (segments #2 and #3). It is also obvious from the Table 4.1b is that we have two candidate line matches over the segment #1, however, visually, it is clear that, segment #1 in the left image corresponds to the segment #a in the right image. To solve the ambiguity, we evaluate the redundancy within the pair-wise matches (Table 4.1a). To be specific, we search the total number of occurrences of each one to one

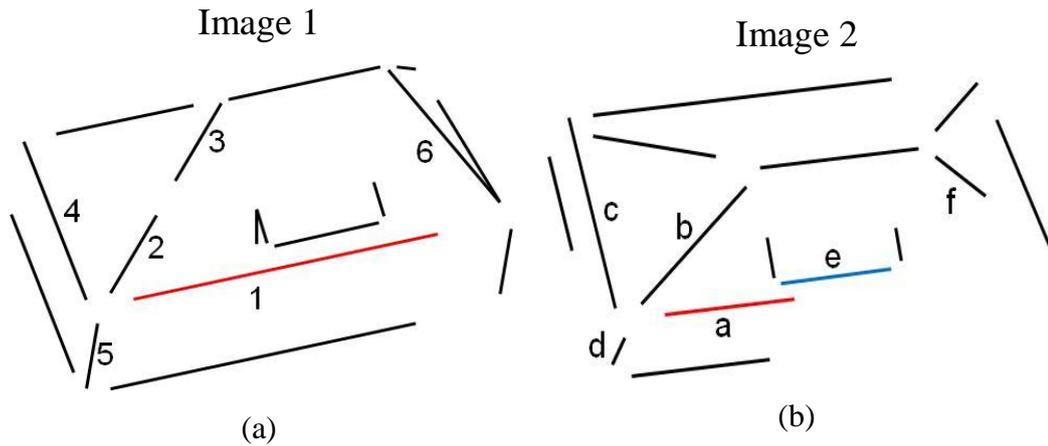


Figure 4.18 Line segments extracted from two stereo images.

Table 4.1 The results of the (a) pair-wise matching and (b) the inferred one-to-one line matches from the pair relations.

Pair-wise Matches	
Left	Right
1 - 2	a - b
1 - 3	a - b
1 - 4	a - c
1 - 5	e - d
1 - 6	a - f

(a)

Line Matches	
Left	Right
1	a - e
2 - 3	b
4	c
5	d
6	f

(b)

relation within the entire pair-wise relations. This gives us quite powerful unique information, since most of the ambiguities occur due to accidental alignments and have a very limited chance to occur in multiple times. For example, in Table 4.1a, among the total of five pair-wise relations that involve segment #1, four out of five corresponds to the segment #a in the right image. Only one pair-wise relation indicates the correspondence with segment #1 and #e for the left and right images, respectively. Thus, the number of occurrences computed from the pair-wise relations over segment #1 reveals the segment #a as the correct match.

In general, the ambiguities can be successfully solved with just counting the number of occurrences. However, due to the perspective distortion, in some rare cases, the number of pair-wise occurrences of wrong candidate may exceed the number of occurrence of the correct match. One good example for this problem is illustrated in Fig. 4.19. It is shown in the figure that, after the pair-wise matching, the line represented with red in the left image have two candidates for matching (two different red lines in the right image). In the next figure, the left column (Fig. 4.20 a, c) and right column (Fig. 4.20 b, d) belong to the left and right stereo images, respectively, and the blue lines demonstrate the lines that assist the pair relations for both candidates. As expected, the correct match (Fig. 4.20 a, b) had successfully paired with a total of four lines that belong to the surrounding boundaries of the building roof. However, surprisingly, the wrong candidate (Fig. 4.20 c, d) paired with a total of six lines (some of them are multiple matches) extracted from the boundaries of a car parked on the nearby street. Thus, for this example, blindly counting the number of occurrences may lead the redundancy measure to a wrong match (red lines in Fig. 4.20 c, d). Therefore, we weight all pair relations proportional to their within pair minimum distances. By this way, the redundancy measure provides a possibility to understand and integrate a local matching support for the line segments. It is clear from Fig. 4.20a and b that

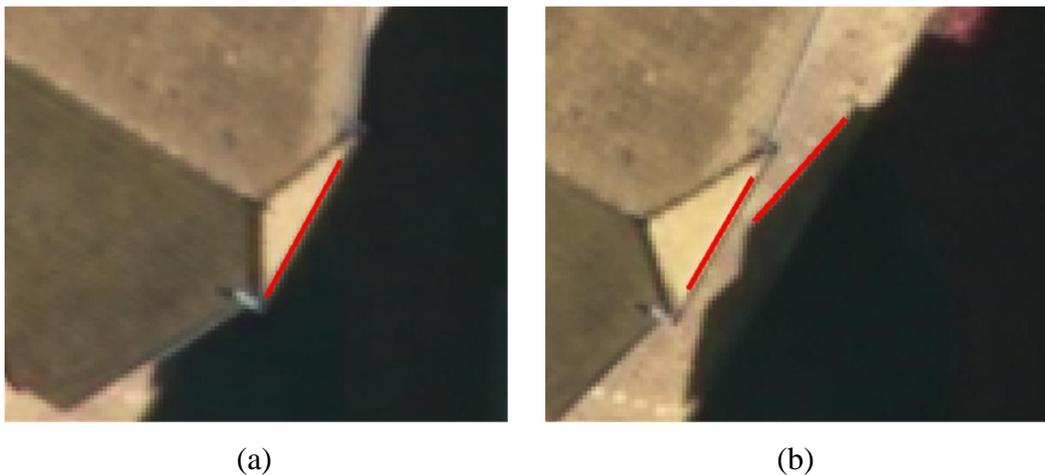


Figure 4.19 Left and right stereo images and one of the matching ambiguities due to the perspective viewing and occlusion.

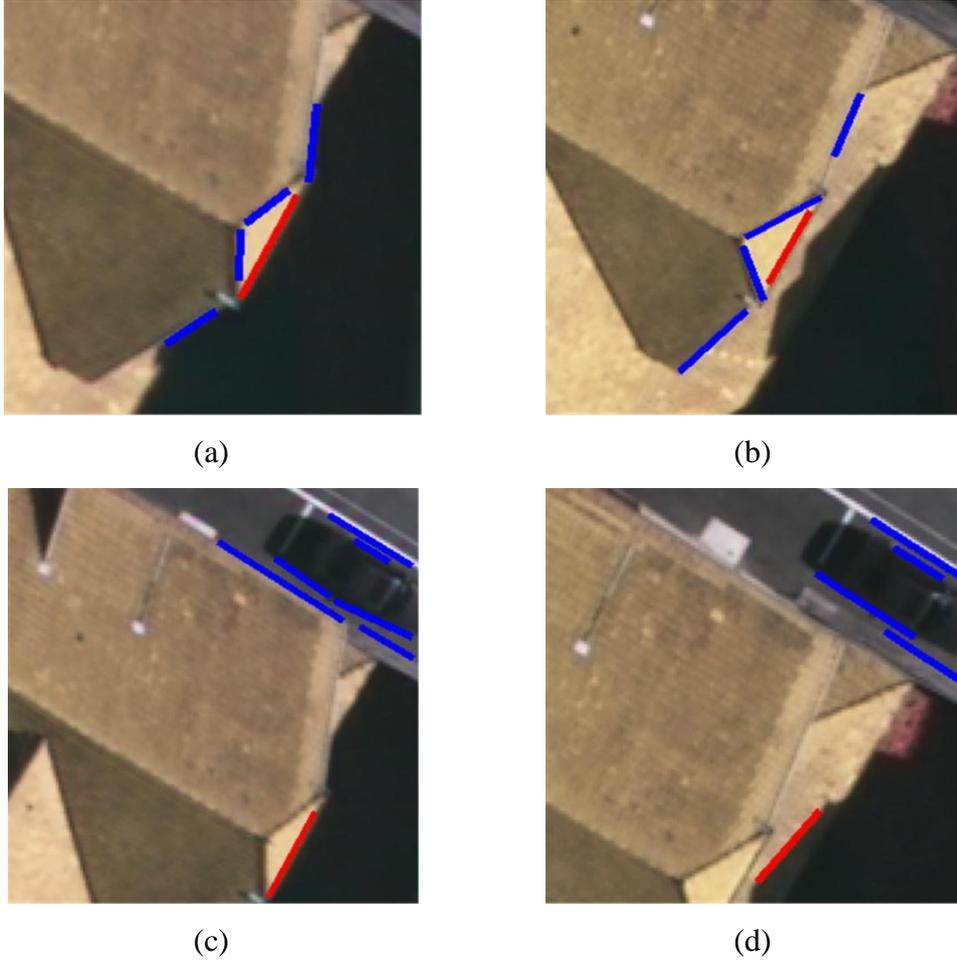


Figure 4.20. A problematic case of the Redundancy measure.

minimum distances between the lines in pair relations that belong to the correct match are much shorter than the ones that belong to the wrong match (Fig. 4.20 c, d). Thus, we propose the new redundancy measure ( $Sim_R$ ) for a line pair as:

$$Sim_R = \sum_{q=1}^N \frac{1}{\sqrt{(d_{ij}^{[L]}, d_{ij}^{[R]})}}, \quad (4.17)$$

provided that the  $d_{ij}^{[L]} \neq 0$  and  $d_{ij}^{[R]} \neq 0$ . In Eq. 6,  $N$  is the number of pair relations assist to matching,  $d_{ij}$  is the pixel-based minimum 2D Euclidean distance between two lines ( $l_i$  and  $l_j$ ) in a pair,  $L$  and  $R$  indicates the pair relations in left and right images, respectively.

### 4.4.3 Pair-wise Quality Measure

During pair-wise matching, the final pair matches are assigned after a weighted pair-wise matching similarity score which is computed over a total of seven measures ( $\xi_1 - \xi_7$ ); an epipolar, three geometric, one photometric, a correlation and a regional measure (see section 4.3.6). The overall pair-wise similarity score ( $\theta_S$ ) for each pair that is computed from those seven measures may give us a hint about the quality of the line matches in that pair. Thus, if a line match is a part of  $N$  number of pairs, the pair-wise quality metric ( $Sim_Q$ ) for that line match is computed as the average of all pair-wise similarities:

$$Sim_Q = \frac{1}{N} \sum_{q=1}^N \theta_S^{[q]}, \quad (4.18)$$

where  $\theta_S^{[q]}$  is the overall pair-wise similarity of the  $q$ th pair relation.

### 4.4.4 Precise Matching

In this study, the measures developed for the precise matching is integrated for the final disambiguation process in a newly developed iterative pair-wise manner, in which the matching inconsistencies are eliminated using nearest/next distance ratios (NNDR) and a final similarity voting scheme. The pseudo code of the developed approach is given below:

- ❖ Eliminate indisputably wrong relations using NNDR computed for the Daisy measure (*ratio<sub>D</sub>*)
- ❖ Initiate Iterative Precise Matching
  - *while ambiguity exist*
  - Compute the Total Similarity Metric (*Sim<sub>T</sub>*)
  - Initiate matching with the highest *Sim<sub>T</sub>* value and label the relation that has the highest value as correct

- Identify the line segments that violate the correct relation
- Check the collinearity of those line segments with the line segments in the correct relation and label them as incorrect if not collinear
- Update pair-wise relation table and update the values ( $Sim_R$  &  $Sim_Q$ )
- *return*

❖ Apply hysteresis Global Thresholding

Our aim during post-processing is to first eliminate indisputably wrong relations based on a very strict NNDR value forced over the measure line-based Daisy ( $Sim_D$ ). NNDR is first introduced by Lowe (2004) based on the fact that the correct matches need to have the closest matching similarity significantly closer than the closest incorrect match to achieve reliable matching. For false matches, there will likely be a number of other false matches within comparable matching similarities (Lowe, 2004). For a large number of datasets, we investigated the NNDR metric in terms of the ratio of closest to second-closest matches of each line (see details in Chapter 6), and for the line matching problem, we reject all the related matches of a match that has a line-based Daisy dissimilarity ( $1 - Sim_D$ ) ratio lower than  $ratio_D$ . The threshold is selected in a way that only a very limited number of line matches that have enough confidence have possibility to fulfill this threshold. Thereafter, we delete the line relations indicated by NNDR from the pair-wise matches, and for each match, we update the redundancy and quality metrics. At this point, it should be pointed out that, if a line relation in a pair is found to be wrong, we do not directly delete the pair, since we don't have any inference (correct or incorrect) for the other match in the pair. The example given in Table 4.1a clarifies this fact. For the fourth pair relation (1-5, e-d), assume that we found that the line match (1-e) is wrong. However, we do not have any information about the other match (5-d) in the pair; thus, we cannot directly label the other relation as incorrect (although it may be in some cases). Therefore, since one of the matches in a pair is labeled as wrong, we update the redundancy ( $Sim_R$ ) and quality measures ( $Sim_Q$ ) of the other match in that pair by eliminating the contribution of that pair from its similarity values. By this way, for example, the

match (5–d) given in the Table 4.1a is not directly eliminated, but penalized, due to reason that the relation (1–e) in the pair is labeled as incorrect.

Once all the measures are updated, we initiate an iterative matching scheme by starting from the match that has the highest total similarity value. For all matches, the total similarity metric ( $Sim_T$ ) is computed by integrating the similarity measures in a weighted linear combination:

$$Sim_T = w_D \cdot Sim_D + w_R \cdot \frac{Sim_R}{(Sim_R)^+} + w_Q \cdot Sim_Q \quad (4.19)$$

In Eq. 4.19, for each selected match, we normalize the Redundancy measure (between 0 and 1) with the maximum Redundancy value of the selected matches, so that the contribution of all similarities is consistent for the final voting. Based on our experiments, we found that the Redundancy is the most reliable and unique measure among the three measures, thus, in this study, weights of the similarities in Eq. 4.19 are designed as  $\{w_D, w_R, w_Q\} = \{1/4, 1/2, 1/4\}$ .

Apparently, among the selected matches, the correct match is the one that maximizes the total similarity metric ( $Sim_T$ ). Thereafter, we fix the correct match and check for the matching ambiguities that violate the selected match. At this point, the collinearity of the line segments of the matching violations (if found any) are individually tested with the line segments of the correct match in order to avoid the deletion of the fragmented lines. The ones that are found to be collinear are labeled along with the correct match for the final matching list. The ones that are not collinear are deleted from the pair-wise matches. After the deletion, we apply the same updating strategy as explained above. Thus, at the end of each iteration, we penalize all related matches in the pairs that are labeled as incorrect. Thus, the (updated) measures ( $Sim_R$  and  $Sim_Q$ ) turn out to be more and more reliable after each iteration.

Finally, the iterations stop after there is no ambiguity left in the final matching list. Like any other system developed so far, if a line segment in the first image has no corresponding line segment in the second image, the system cannot identify and label the match as incorrect (if accidentally assigned) since the correct line to be matched is missing. To solve this problem, a final check with a global threshold is required. On the contrary to the previous studies that rely on a single threshold, we propose a new hysteresis-like global thresholding to solve the problem and to maximize the performance of the matching. As we penalize the Redundancy measure ( $Sim_R$ ) for each match after each iteration, once the iterations has stopped, we have a near-perfect final ( $Sim_R$ ) values for the final matching list. This gives us a unique way to solve the above mentioned problem, in principle; those ill-posed matches have very low Redundancy values when compared to values of the correct matches. Thus, we define a two-level global thresholding:

- (i).  $Sim_D \geq Thr_D^1$
- (ii).  $Sim_D \geq Thr_D^2 \quad \& \quad Sim_R \geq Thr_R$

From experiments, we have found that a global Daisy threshold of  $Thr_D^1$  must be independently satisfied by every match. However, due to lack of rich textures in local line neighborhood, some false matches may easily exceed this threshold. Increasing the threshold may have a possibility to eliminate some of the correct matches as well, thus results in reducing the overall completeness of the matching. So, we propose to utilize a second high Daisy threshold,  $Thr_D^2$  restricted with a global Redundancy threshold of  $Thr_R$ . By this way, compared to case where only a single global threshold is enforced, using a two-level thresholding at the same time can eliminate most of the remaining false matches while keeping the matching precision and matching completeness.

## CHAPTER 5

### STEREO LINE RECONSTRUCTION

In this chapter, our motivation and contribution is mainly on the height estimation of the matched segments. Although the method of direct construction gives satisfactory reconstruction results for the lines that are not aligned with the epipolar line, a dramatic decrease in terms of height accuracy for the lines that are nearly or exactly aligned ( $\leq 10^\circ$ ) with the epipolar line is inevitable (Fig. 5.1). This is due to the reason that if the angles of lines in image space get closer to the epipolar direction, the two projection planes generated from line segments become similar and in the worst case (exact alignment) they turn out to be the same plane. For those cases, the direct construction of 3D lines from the intersection of planes is highly problematic in terms of final height accuracy and in some cases the intersection (or the reconstruction) may not be possible. Therefore, in this thesis, a new reconstruction method which relies on the relations developed in the pair-wise approach is developed. The main idea is to manipulate the redundancy inherent in pair-relations to generate artificial 3D point entities ( $\mathbf{X}_i$ ) from available pair matches and utilize those points during the estimation process to improve the height estimation of the matched segments. However, since we do not exactly know whether the two lines in a pair really intersect on the Earth surface or not, before the estimation process, we select the proper point entities by means of a new weight function which is composed of mainly three terms computed in a pair-wise manner: Euclidean distance, epipolar constraint and intersection angle in image space. For each problematic matching case ( $\leq 10^\circ$ ), we automatically select the appropriate artificial 3D point entities and

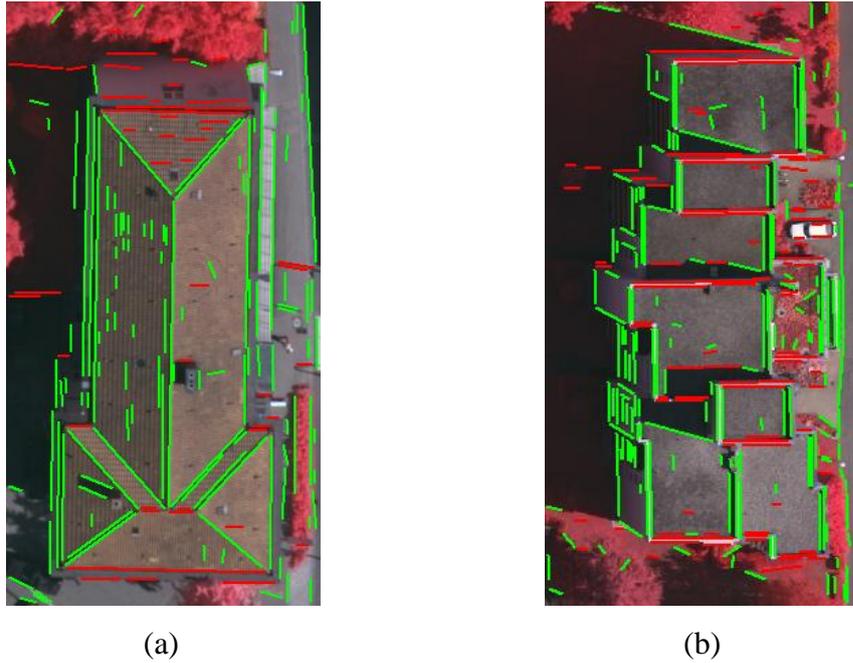


Figure 5.1 The line segments extracted from two images. The red ones indicate the lines that are aligned ( $\leq 10^\circ$ ) with the epipolar line (flight direction).

integrate each selected entity during the estimation process along with the projection planes of the problematic line segments. Thus, at the end of this joint estimation, we have a possibility to reconstruct those problematic line segments with promising final performances.

### 5.1 Reconstruction of Line Segments

It is well-known that the reconstruction of straight lines which are nearly parallel to the epipolar line is numerically unstable within a stereo image pair or a single image strip (Zhang, 2005). For that reason, Zhang (2005) proposed an alternative way which relies on free-form line structures generated from multiple line segments. The major difficulty with the free-form structures is that the problematic line(s) must have an edge connection with the other neighboring lines; however, this is rarely the case for most of the line segments. Here, we present a new approach which also relies on line to line relations developed in the pair-wise approach. Thus, we have a possibility to reconstruct those problematic

line segments without having such an edge connectivity assumption.

The reconstruction process starts with a test which determines the angle difference ( $0 - 90^\circ$ ) between the line segments and the related epipolar line. Based on our experiences, the line segments that have angle differences of less than 10 degrees are highly susceptible to produce inaccurate reconstruction results. Therefore, during the test, we label the lines as not-aligned if the angle difference is computed to be larger than 10 degrees. The reconstruction of all those matched line segments are performed by intersecting the projection planes (Fig. 5.2),  $\mathbf{A}^1(l_1)$  and  $\mathbf{A}^2(l_2)$ , with the method of direct construction as introduced in Heuel and Förstner (2001) and Heuel (2001).

The underlying approach for the reconstruction of line segments that is nearly-aligned ( $\leq 10^\circ$ ) with the epipolar line is to manipulate the redundancy (see Chapter 4, section 4.4.2) inherent in pair-relations to generate artificial 3D point entities ( $\mathbf{X}_i$ ) and utilize those points during the estimation process. By this way, the neighbouring line segments that have a pair-connection with the problematic segment contribute to the height estimation (Fig. 5.3). However, the reconstruction cannot be performed in a single framework, since all 2D

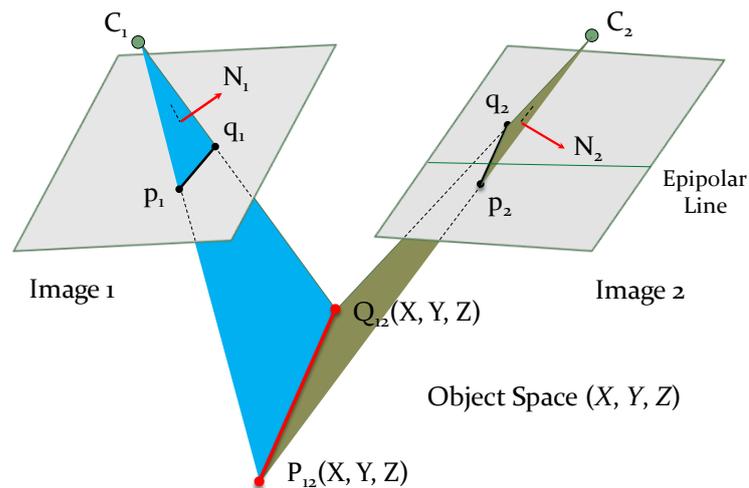


Figure 5.2 The reconstruction of line segments that are not aligned with the epipolar line.

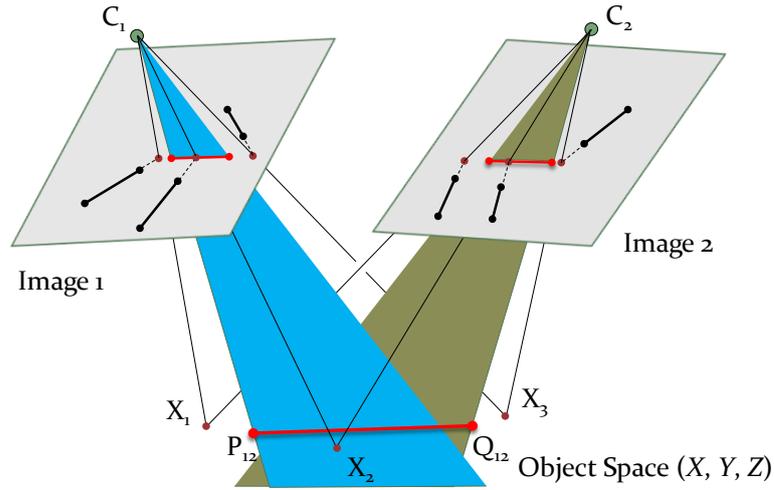


Figure 5.3 The reconstruction of line segments that are nearly-aligned with the epipolar line.

intersection points generated lie exactly on the problematic line segments; thus also belong to the projection planes. Therefore, each artificial 3D point entity ( $\mathbf{X}_i$ ) must be generated beforehand, and the final estimation should be jointly performed along with the related projection planes ( $\mathbf{A}_i$ ). In this section, we follow the similar representation of the homogeneous uncertain vectors given in Heuel and Förstner (2001).

### 5.1.1 The Generation of Uncertain Artificial 3D Point Entities:

Formally the uncertain homogeneous vectors are denoted as  $(\mathbf{x}, \Sigma_{xx})$ , where  $\Sigma_{xx}$  is the covariance matrix of the homogeneous vector  $\mathbf{x}$ . A 2D uncertain line  $(\mathbf{l}, \Sigma_{ll})$  in image space can be generated by joining the end-points of the line segment,  $(\mathbf{x}, \Sigma_{xx})$  and  $(\mathbf{y}, \Sigma_{yy})$ :

$$(\mathbf{l}, \Sigma_{ll}) = (\mathbf{x} \times \mathbf{y}, S(\mathbf{y}) \Sigma_{xx} S(\mathbf{y})^T + S(\mathbf{x}) \Sigma_{yy} S(\mathbf{x})^T) \quad (5.1)$$

where skew-symmetric matrix  $S(\mathbf{x})$  and  $S(\mathbf{y})$  are the matrix representation of the points  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. Similar to Eq. 1, the intersection point  $\mathbf{x}$  of two lines  $\mathbf{l}$  and  $\mathbf{m}$  can be computed as

$$(\mathbf{x}, \Sigma_{xx}) = (\mathbf{l} \times \mathbf{m}, S(\mathbf{l}) \Sigma_{mm} S(\mathbf{l})^T + S(\mathbf{m}) \Sigma_{ll} S(\mathbf{m})^T) \quad (5.2)$$

with the related skew-symmetric matrices,  $S(\mathbf{l})$  and  $S(\mathbf{m})$ . In our case, initially, the covariance matrices of each line endpoints are computed from edges that form the line segment, the details are given in Chapter 3. Next, the lines and their uncertainties are computed from those endpoints using Eq. 5.2 (Fig. 5.4a). Thereafter, for each problematic segment, the neighbouring line segments that have a pair-connection with the problematic segment are collected (from the available pair-wise matches) and their intersection points are computed using Eq. 5.2. Fig. 5.4b illustrates an example of a generated artificial intersection point with its elliptical confidence region.

The estimation of the artificial 3D point entities ( $\mathbf{X}_i$ ) from 2D correspondences cannot be performed using direct construction since the projecting rays emerge from 2D points rarely intersect in object space. Therefore, we estimated the 3D points using the iterative linear estimation model Gauss-Helmert with constraints.

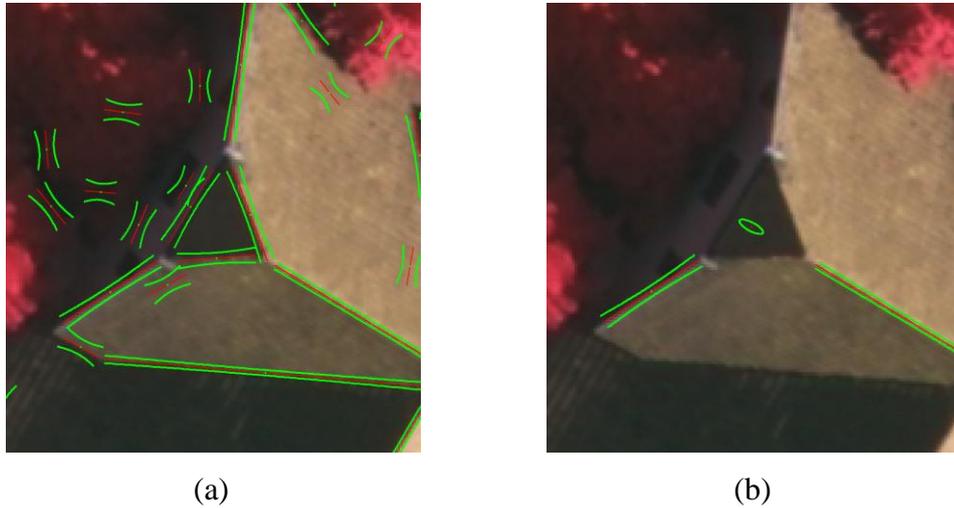


Figure 5.4 (a) Confidence regions of the extracted line segments, (b) an example of the generated artificial intersection point and its elliptical confidence region. Entities and confidence regions are shown in red and green colors, respectively. Confidence regions are 15 times exaggerated for visualization.

The details of the Gauss-Helmert model can be found in Appendix A. In our case, we have 6 observations for each case (two homogeneous 3 vectors for image points  $\mathbf{x}_{ij}$ ) and 4 unknowns for the homogeneous coordinates of each 3D point entity. Thus, referring to the Gauss-Helmert model, we have the vector  $\mathbf{l}$  of the observations and the vector  $\mathbf{u}$  of the unknown parameters:

$$\mathbf{l}_{6 \times 1} = \begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{j2} \end{bmatrix} \quad \mathbf{u}_{4 \times 1} = \mathbf{X}_i \quad (5.3)$$

With the aid of the reduced skew-symmetric matrices  $S^{[r]}(\cdot)$  (Manual of Photogrammetry, 2004), there are 2 independent constraints  $\mathbf{g}$  for each observed image point  $\mathbf{x}$  and the unknown parameters ( $\mathbf{X}_i$ ). In addition, we have a single length constraint  $h$  on the unknown parameters. Thus, the relations of the Gauss-Helmert model for the estimation of the 3D point entities can be written as:

$$\mathbf{g}(\mathbf{l}, \mathbf{u})_{4 \times 1} = \begin{bmatrix} S^{[r]}(\mathbf{x}_{i1})P_1\mathbf{X}_i \\ S^{[r]}(\mathbf{x}_{j2})P_2\mathbf{X}_i \end{bmatrix} = \mathbf{0} \quad (5.4)$$

$$h(\mathbf{u}) = \mathbf{X}_i^T \mathbf{X}_i - 1 = 0$$

where  $P_1$  and  $P_2$  are the 3x4 projection matrices for points on images 1 and 2, respectively. The initial approximate values of  $\mathbf{X}_i$  for the iterative solution can be obtained from the SVD solution (Heuel and Förstner, 2001), and once the estimation is completed, the covariance matrices of the estimated 3D point entities can be computed from the inverted normal equation matrix.

### 5.1.2 The Joint Estimation of the 3D Line Segments:

For the estimation procedure, we parameterize the 3D lines in Plücker representation  $\mathbf{L}^T = (\mathbf{L}_h^T, \mathbf{L}_0^T) = (L_1, L_2, L_3; L_4, L_5, L_6)$  and utilize an iterative linear Gauss-Markoff model with constraints (Appendix A). The algebraic expressions of the form  $\mathbf{g}_i(\beta; \gamma_i) = \mathbf{0}$  with respect to all possible observations ( $\gamma_i$ ) and unknown ( $\beta$ ) entities are developed and explicitly given in Förstner et. al.

(2000) and Heuel and Förstner (2001). In our case, we are searching for an unknown line  $\mathbf{M}$  which primarily must lie in two planes; thus,  $\mathbf{g}_1(\mathbf{M}; \mathbf{A}_i) = \Pi^T(\mathbf{A}_i)\mathbf{M} = \mathbf{0}$  where  $\Pi$  is the homogeneous matrix representation of planes  $\mathbf{A}_i$ . The projection planes  $\mathbf{A}_i$  for each line  $\mathbf{l}_i$  and the related uncertainties can be determined using the projection matrices  $\mathbf{P}_j$  of the images (Förstner, 2010):

$$\begin{aligned} \mathbf{A}_i &= \mathbf{P}_j^T \mathbf{l}_i \\ \Sigma_{\mathbf{A}_i \mathbf{A}_i} &= \mathbf{P}_j^T \Sigma_{\mathbf{l}_i \mathbf{l}_i} \mathbf{P}_j + (\mathbf{I}_4 \otimes \mathbf{l}_i) \Sigma_{\mathbf{P}_j \mathbf{P}_j} (\mathbf{I}_4 \otimes \mathbf{l}_i^T) \end{aligned} \quad (5.5)$$

where,  $\mathbf{I}_n$  represents  $n \times n$  unit matrix and  $\otimes$  denotes the Kronecker product. The uncertainty of each line is derived from Eq. 5.1 and for this study we assume that the projection matrices are free of error. In addition to the projection planes, the unknown line must also satisfy the artificial 3D point entities ( $\mathbf{X}_i$ ) generated from the neighbouring line segments,  $\mathbf{g}_2(\mathbf{M}; \mathbf{X}_i) = \bar{\Pi}^T(\mathbf{X}_i)\mathbf{M} = \mathbf{0}$ , where  $\bar{\Pi}$  is the homogeneous matrix form of points  $\mathbf{X}_i$ . However, since we do not exactly know whether the two lines in a pair really intersect on the Earth surface or not, before the estimation process we compute weights for each 3D point entity generated. The weights utilized depend on three measures computed in a pair-wise manner; minimum 2D Euclidean distance ( $d_{ij}$ ) between two lines ( $l_i$  and  $l_j$ ), the minimum angle ( $\theta_{ij}$ ) enclosed by line segments  $l_i$  and  $l_j$ , and the minimum orthogonal distance ( $d_{ij}^e$ ) between the intersection points and related epipolar lines ( $l_{epi}$ ). In principle, the reliability of a point increases if the distances (both  $d_{ij}$  and  $d_{ij}^e$ ) computed are relatively short and decreases if the intersection angle is quite narrow (ex.  $< 10^\circ$ ). Therefore, we designed the new cumulative weight function ( $W_i$ ) as:

$$\begin{aligned} W_i &= t_{ij} \cdot e^{-\left(\frac{\sigma_2 d_{ij} + \sigma_1 d_{ij}^e}{2\sigma_1\sigma_2}\right)} \\ t_{ij} &= \begin{cases} 0 & \text{if } \theta_{ij} \leq 10^\circ \\ 1 & \text{if } \theta_{ij} > 10^\circ \end{cases} \end{aligned} \quad (5.6)$$

where the parameter  $\sigma_1$  and  $\sigma_2$  controls the weighting for metrics  $d_{ij}$  and  $d_{ij}^e$ , respectively. Fig. 5.5 shows the weighting curves separately computed for each measure.

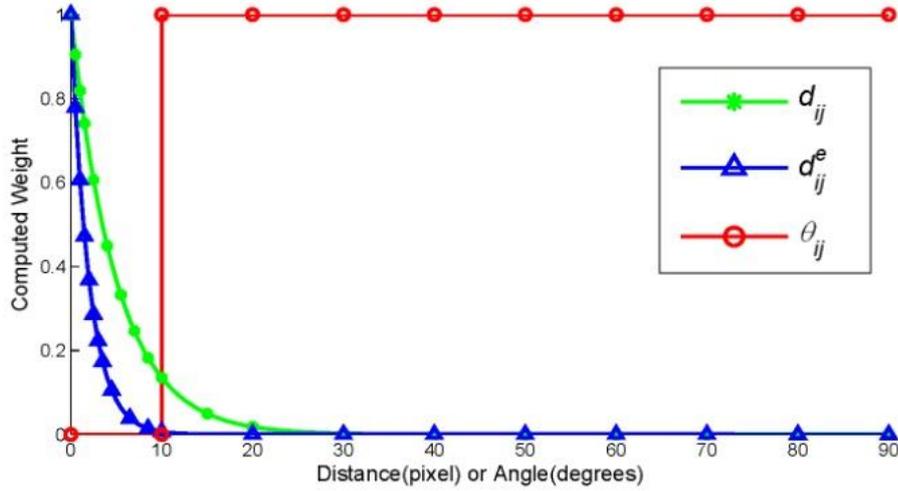


Figure 5.5 Independent weight curves for each measure. The control parameters ( $\sigma_1$  and  $\sigma_2$ ) are (5 and 2), respectively.

For each nearly-aligned matching case ( $\leq 10^\circ$ ), the existing 3D point entities ( $\mathbf{X}_i$ ) are collected and their weights ( $W_i$ ) are automatically determined using Eq. 5.6. However, it is not logical to integrate all observed point entities directly to the estimation process, since some of those points may be generated from wrong matches. Therefore, first, we eliminate all point entities that have weights less than a pre-defined threshold ( $T_w \leq 0.05$ ). Thereafter, among the remaining point entities, only the points that have the highest weights are integrated to the estimation process along with the observed projection planes. However, the selection approach for the best point entities is not trivial and must be carefully handled. Fig. 5.6 demonstrates a case in which a matching line ( $l_e$ ) is aligned with the epipolar line. Assume that  $l_e$  is paired with three neighbouring lines ( $l_1$ ,  $l_2$ , and  $l_3$ ) after matching. We estimate the related 3D point entities ( $\mathbf{X}_1$ ,  $\mathbf{X}_2$ , and  $\mathbf{X}_3$ ) from each corresponding intersection point ( $p_{ij}$ ) using the Eqs. 5.1-5.4 and compute the related weights ( $W_1$ ,  $W_2$  and  $W_3$ ) for each entity from Eq. 5.6. However, it is clear

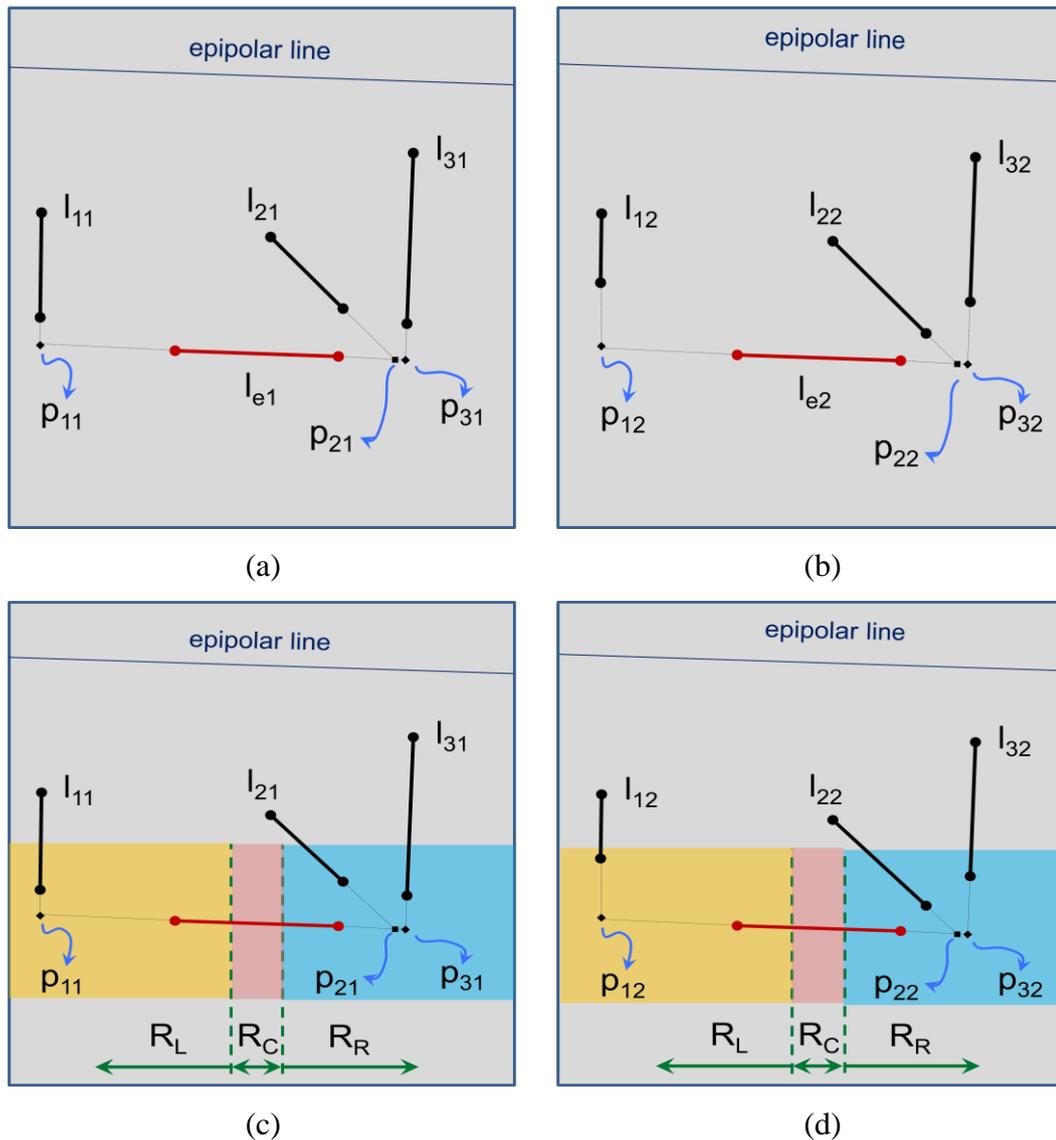


Figure 5.6 Left (a-c) and right (b-d) stereo images. The directional regions utilized during the selection of the best weighted point entities.

that the entities ( $\mathbf{X}_2$  and  $\mathbf{X}_3$ ) will get higher weights than entity  $\mathbf{X}_1$ , since the computed minimum 2D Euclidean distance ( $d_{ij}$ ) measure for  $\mathbf{X}_1$  is significantly larger than the other distances. Thus, for this case, the 3D line will be estimated from the entities ( $\mathbf{X}_2$  and  $\mathbf{X}_3$ ) along with the projection planes ( $\mathbf{A}_{e1}$  and  $\mathbf{A}_{e2}$ ). However, since those entities are very close to each other and are located just one side (right) of the lines ( $l_{e1}$  and  $l_{e2}$ ) the final position of the line segments in object space is highly sensitive to the small deviations between the entities  $\mathbf{X}_2$  and  $\mathbf{X}_3$ ,

which may significantly reduce the final accuracy of the 3D line estimated. To solve this problem, we propose a region based weighting of points (Fig. 5.6c-d) in which the highest weighted points are calculated for each region separately. We utilize two split points ( $l/3, 2l/3$ ) to divide the direction dominated by each line into three classes, left ( $R_L$ ) – center ( $R_C$ ) – right ( $R_R$ ). Next, best weights are computed for each region separately. Thus, we guarantee that the point entity ( $X_1$ ) generated by the intersection points ( $p_{11}$  and  $p_{12}$ ) contributes to the final estimation.

For the estimation process, we form the point-line and plane-line incidence relations in 3D space and perform an iterative linear Gauss-Markoff model with constraints. For each case, we have six unknowns and two constraints (Plucker and length) for each 3D line ( $L_i$ ). In addition to the two projection planes, the final observation number depends on the number of point entities  $X_i$  available for each region. Once again, the required initial values can be taken from the SVD solution and the covariance matrices of the estimated 3D line entities can be computed from the inverted normal equation matrix of the Gauss-Markoff model.

## **CHAPTER 6**

### **DISCUSSION: TEST DATA AND PERFORMANCE EVALUATION**

This chapter presents the performance evaluation of the proposed stereo line matching and reconstruction approach. The assessments are performed for both analogue and digital images acquired over two different parts of Germany. To evaluate the results of the line matching, the quantitative line matching results in terms of the correctness, completeness and quality levels are provided. The performance of the reconstruction is assessed for three image patches selected from the test site Vaihingen (Cramer, 2010) by comparing them with the highly accurate reference data obtained from airborne Light Detection and Ranging (LIDAR).

#### **6.1 Test Sites and Image Datasets**

##### **6.1.1 The Vaihingen/Enz Test Site**

The Vaihingen/Enz test site was first established for the geometrical performance tests of one of the first digital airborne line scanning systems in 1995. After that, the test site was used several times for different kinds of investigations: For the independent in-flight evaluation of new digital airborne sensors as well as for investigations on the potential of direct georeferencing using integrated GPS/inertial systems in combination with standard analogue frame cameras (Web 1). The test site is located about 20km north-west of the city of Stuttgart-Germany and the spatial extension of the overall test area is about 7.5 km (east-west) x 4.7

km (north-south). The terrain heights differ between 193m and 356m above the mean sea level. The terrain variations, the extent of the overall test site, and the image area utilized from the test site are depicted in the Figure 6.1 (vertical component 2x exaggerated, © Landesvermessungsamt Baden-Württemberg) (Web 1).

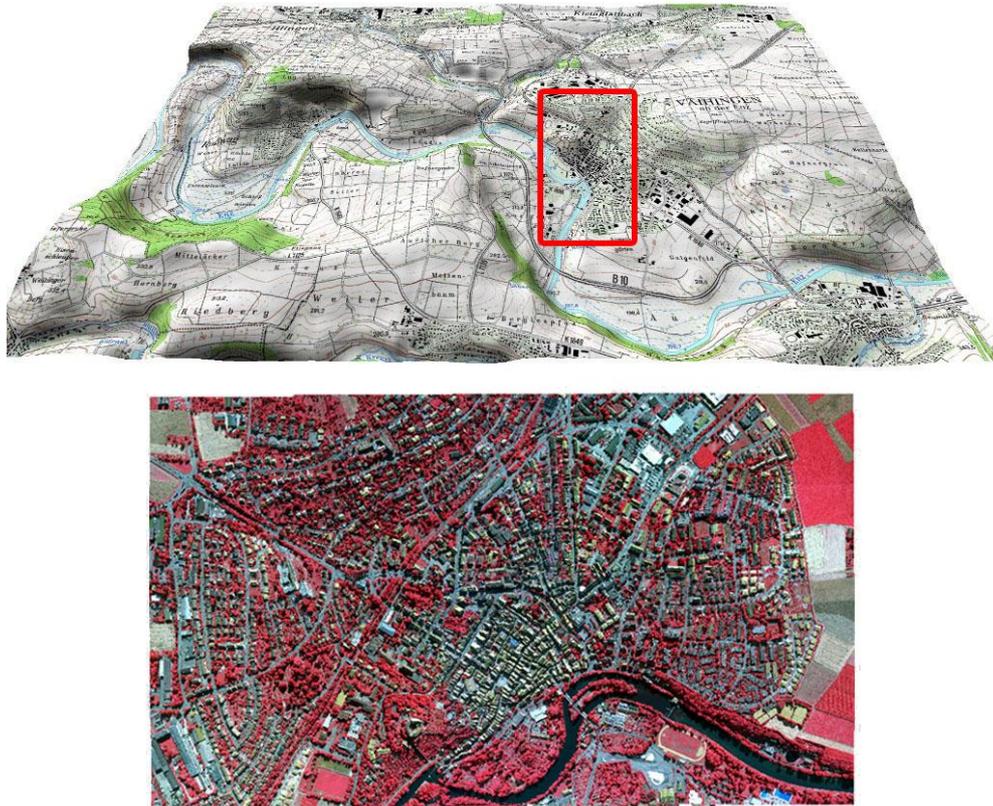


Figure 6.1 Vaihingen/Enz test site (Web 1) and the aerial ortho-image (rotated 90° CCW)

From the overall test site, nine UHR Digital Mapping Camera (DMC) images from three different strips covering the urban area of the city of Vaihingen are selected. The images were acquired with 70% forward overlaps with a base-to-height (B/H) ratio of 0.28. The focal length of the camera was 120 mm and the flying heights for the test fields were approximately 800 m above ground, which corresponds to ground sampling distances (GSD) of approximately 8 cm. The bundle adjustment of the test site was reported as sub-pixel level in planimetry

and a pixel level in elevation. Additional flights with almost a month time difference with the DMC acquisitions were done with a Leica ALS 50 LIDAR sensor in order to produce a highly accurate reference height data of the test site. The accuracy evaluations of the LIDAR data proved that the acquired LIDAR point cloud has an overall height accuracy of 3.3 cm (Haala et al., 2010).

### 6.1.2 The Hannover Test Site

The city of Hannover (Fig. 6.2) is the capital of the federal state of Lower Saxony (Niedersachsen) of Germany which is the second largest in area and the fourth crowded in population among the sixteen states of Germany. The selected test site is located near to the central part of the Hannover city with a spatial extent of

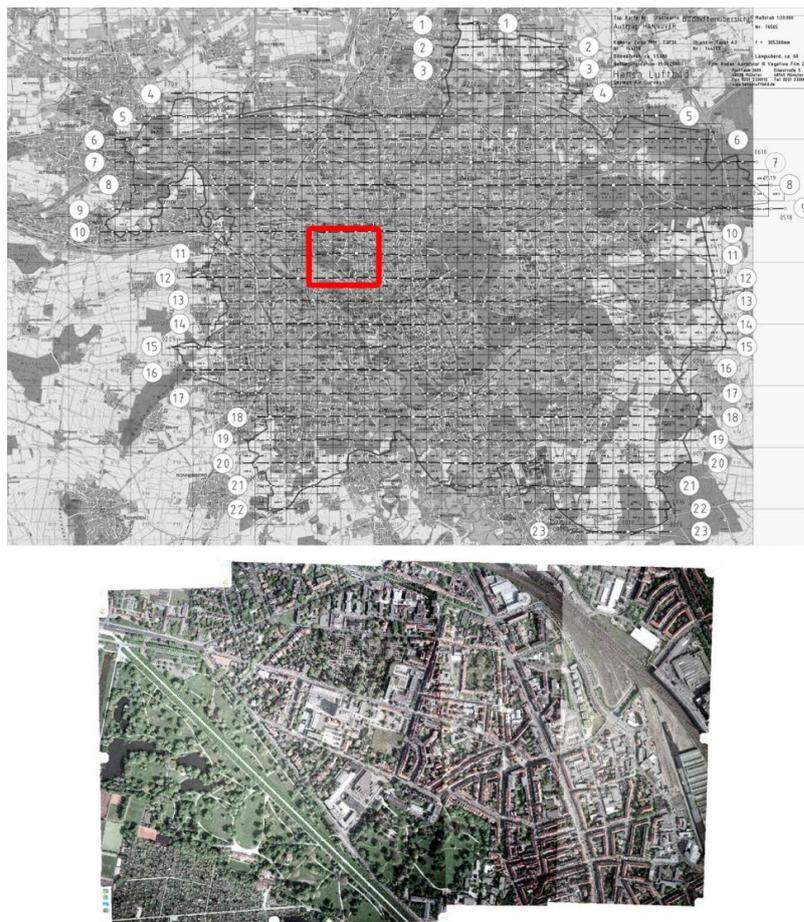


Figure 6.2 The Hannover test site and the aerial ortho-images

approximately 1.3 km (east-west) x 2.1 km (north-south) (© Geoinformation Hannover). The site mainly consists of built-up and residential environments with very dense and complex buildings including other manmade objects. The site has a flat topography, with an average elevation of 54 m.

The UHR aerial images of the city of Hannover were taken with Zeiss RMK TOP30 analog camera with standard 60% forward overlaps with a B/H ratio of 0.3. The calibrated focal length of the camera was 305.560 mm with a flying height of approximately 1600 m. The images were scanned at 14  $\mu\text{m}$  resolution and this corresponds to a final GSD of approximately 7 cm. The bundle adjustment of the test site was performed as a sub-pixel level both in planimetry and in elevation.

## **6.2 Assessment Strategies for the Line Matching and Reconstruction**

### **6.2.1 Assessment Strategy for the Line Matching**

In order to assess the accuracy of the line matching, the line matches were classified into three categories: True Positive (TP), False Positive (FP), and False Negative (FN), by comparing the automatically matched line segments with the manually generated reference line matching list. We term a line match as a True Positive if that match correctly corresponds to a matching relation in the reference list. On the other hand, a False Positive is a line match that does not correspond to any of the line matches in the reference list, and a False Negative is a line match that exists in the reference list but cannot be found by the automated approach. In this thesis, we follow the well-known three metrics to evaluate the quality of the line matching (Rutzinger et. al., 2009):

$$\textit{Correctness} = \frac{\|TP\|}{\|TP\| + \|FP\|} \quad (6.1)$$

$$Completeness = \frac{\|TP\|}{\|TP\| + \|FN\|} \quad (6.2)$$

$$Quality = \frac{\|TP\|}{\|TP\| + \|FP\| + \|FN\|} \quad (6.3)$$

where  $\|.\|$  denotes the number of line matches assigned to each specific class. The Quality metric in Eq. 6.3 gives an overview of the performance of the line matching, since a successful matching output must have high rates of completeness and correctness levels at the same time.

### 6.2.2 Assessment Strategy for the Line Reconstruction

For the Vaihingen test dataset, the accuracy of the reconstructed line segments could be evaluated by comparing them to reference LIDAR data. In order to compare the reconstructed line segments, we automatically extracted 3D planes from the point cloud in the vicinity of each line. Depending on the type of the line segment, this plane reconstruction process results in one plane if the line corresponds to a step edge and in two planes if the line corresponds to the intersection of two different planes (Fig. 6.3). Thereafter, for each of the line segments, we determined the line's average orthogonal distance from its

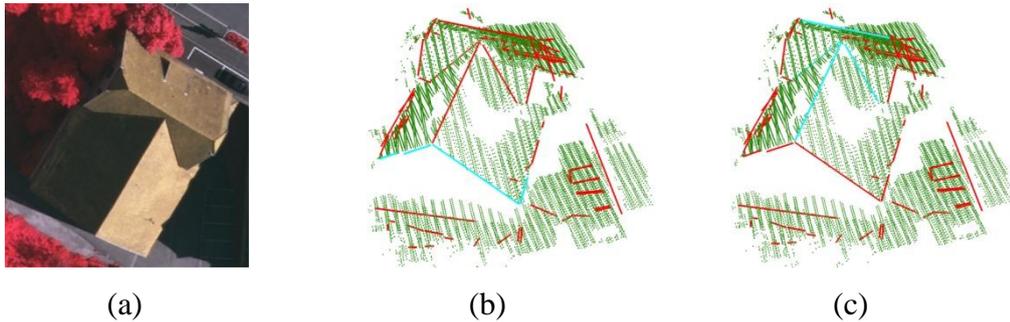


Figure 6.3 (a) A test building, and two different types of line segments observed. Blue segments illustrate some examples for (b) the step edges, and (c) the intersection edges. The reference LIDAR point cloud is shown in green color.

neighbouring planes and used these distances to compute the RMS average distance between the reconstructed line segments and the LIDAR planes. The RMS distance was determined separately for line segments corresponding to one plane and those corresponding to two; furthermore, a total RMS distance was also determined.

### 6.3 The Results and Discussion of the Proposed Line Matching Approach

#### 6.3.1 The Results and Discussion of the Vaihingen Test Site

##### 6.3.1.1 Parameter Selection for the Initial Pair-wise Matching Stage

Both the success and the complexity of the matching are first determined by the proximity distance parameter ( $T_{\text{prox}}$ , see Chapter 4, section 4.1). The parameter controls the total number of pair-wise line relations generated for the matching. Figure 6.4 clearly demonstrates the underlying complexity for various proximity values defined on object space. The figure illustrates the level of increase in the total number of pair-wise relations generated between the line segments for each test patch given in Figure 6.5. It is clear from the figure that the number of

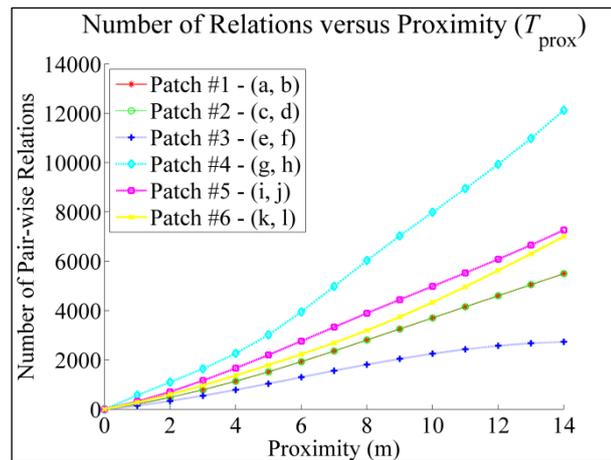


Figure 6.4 The total number of pair-wise relations computed with respect to Proximity measure ( $T_{\text{prox}}$ )

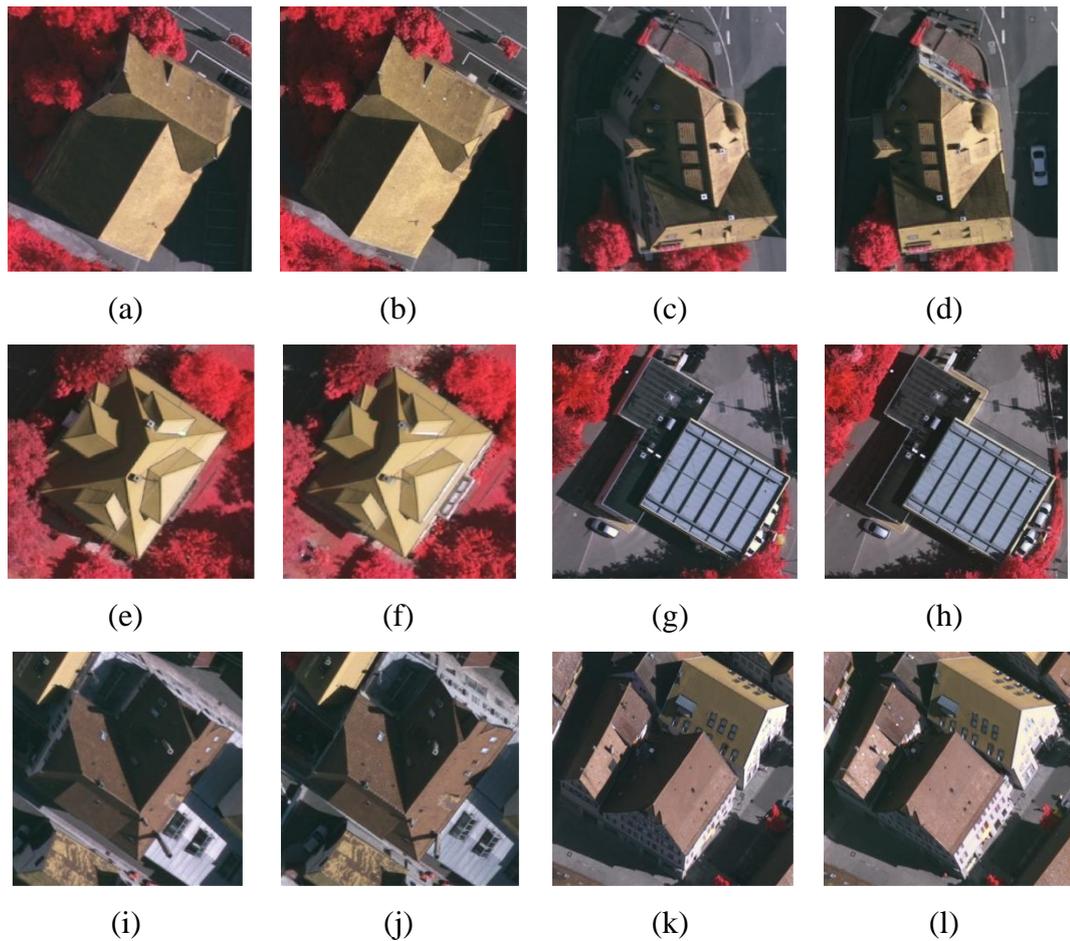


Figure 6.5 Six stereo test patches selected from the Vaihingen test site.

pair-wise relations may reach over ten thousand even for such small urban patches. For sure, the slope of each increasing curves in Figure 6.4 not only depend on the proximity parameter, but also the total number of line segments extracted for each patch. Thus, it is possible to conclude that the highest number of line segments is extracted for the image patch #4 which has the steepest slope among all the six image patches.

Since the pair-wise matching stage initially determines the line segments that will be utilized for the final precise matching stage, it is essential to include all possible correctly matching line segments into pair-wise relations. Thus, although reducing the proximity values to less than 5 m seems to be effective to limit the number of pair-wise combinations generated, short proximity values may prevent

some of the correct line matches from being paired with other nearby line segments. On the other hand, increasing the proximity values to over 10 m ensures that almost all of the correct matches are successfully paired with nearby line segments, but at the same time, the numbers of pairs generated reach an excessive level (Figure 6.4). At this point, manual investigations revealed that a proximity distance of 8 m gives a good sense of balance between the number of pairs generated and the number of correct matches paired successfully. Thus, in this study, the proximity distance is fixed to 8 m on object space for the pair-wise matching of all test patches.

During the initial pair-wise matching stage, various threshold values can be defined for each of the seven pair-wise constraints ( $\zeta_1 - \zeta_7$ ) to further limit the number of incorrect matches. Besides, restricting some of the practically impossible pair-wise matches with thresholds may also have a massive impact on the speed of the processing, since if one constraint is not fulfilled, in that case, there is no need to further test the remaining constraints. However, at this point, as already stated in Kim and Nevatia (2004), the major difficulty arises in the stage of gathering supervised ground truth examples of correct pair-wise matches, because it is rather time consuming to manually assign or provide ground truth for thousands of pair-wise combinations. Therefore, here, a different approach is utilized for this purpose. First, for each constraint, very relaxed thresholds which can hardly be exceeded by the correct pair-wise matches are manually determined. Next, the initial pair-wise matching stage is performed automatically by restricting the matches with those very relaxed thresholds. After that, matching pairs found from the automatic matching results are collected and assessed to determine the values of the thresholds. During the assessments, since both correct and incorrect pair-wise matches are labeled, probability density functions (PDF) for correct pair-wise matches and as well as incorrect matches can be computed. Thus, this also gives a chance to analyze the empirical distribution of the incorrect pair-wise matches to further understand the limits of the initial pair-wise matching stage.

Fig. 6.6 shows the computed PDFs of the correct and incorrect pair-wise matches that are generated using the six image pairs given in Fig. 6.5. For the first two geometrical constraints ( $\zeta_1$  and  $\zeta_2$ ), it is clear that almost 60% of the correct pair-wise relations in stereo images found to be having no difference in terms of their geometrical pair-wise intersection and directional angle values (Fig. 6.6a, b). The maximum threshold values for those constraints seem to be around 30 degrees. For the third geometrical pair-wise constraint ( $\zeta_3$ ), a ratio difference threshold of 0.4 can effectively hold more than 99% of the correct matches. The PDF of the correct pair-wise matches of the radiometric constraint ( $\zeta_4$ ) shows that approximately 60% of the correct pair-wise matches have intra-pair radiometric value differences of 5 or less. It is also clear that almost 95% of the correct matches have pair-wise radiometric differences of less than 25. In terms of the correlation constraint ( $\zeta_5$ ), unsurprisingly, almost 30% of the correct matches have correlation values of less than 0.6. Obviously, this kind of result is related to the violations that may occur during the plane formation (see details in Chapter 4, section 4.3.3). Therefore, a relaxed correlation threshold, such as 0.2, must be selected to cover most of the correct pair-wise matches. The PDF computed for the correct pair-wise matches of regional constraint ( $\zeta_6$ ) shows that a regional similarity threshold of 0.4 can effectively handle more than 95% of the correct matches. For the pair-wise epipolar constraint ( $\zeta_6$ ), we see that about 95% of the pair-wise intersections have orthogonal distance errors of less than 5 pixels, which can be considered as the allowed maximum orthogonal distance error for the Vaihingen test patches. Thus, for the Vaihingen test site, we fixed the pair-wise thresholds as given in Table 6.1.

For all constraints, if we analyze the PDFs of the incorrect pair-wise matches, it is clear that, we do not have very distinct separation between the correct and incorrect pair-wise matches. Fig 6.6a-c shows that, in terms of geometric constraints ( $\zeta_{1-3}$ ), the remaining incorrect pair-wise matches after applying the thresholds have a very similar distribution compared to the distribution of the correct pair-wise matches. Thus, it is almost impossible to differentiate the

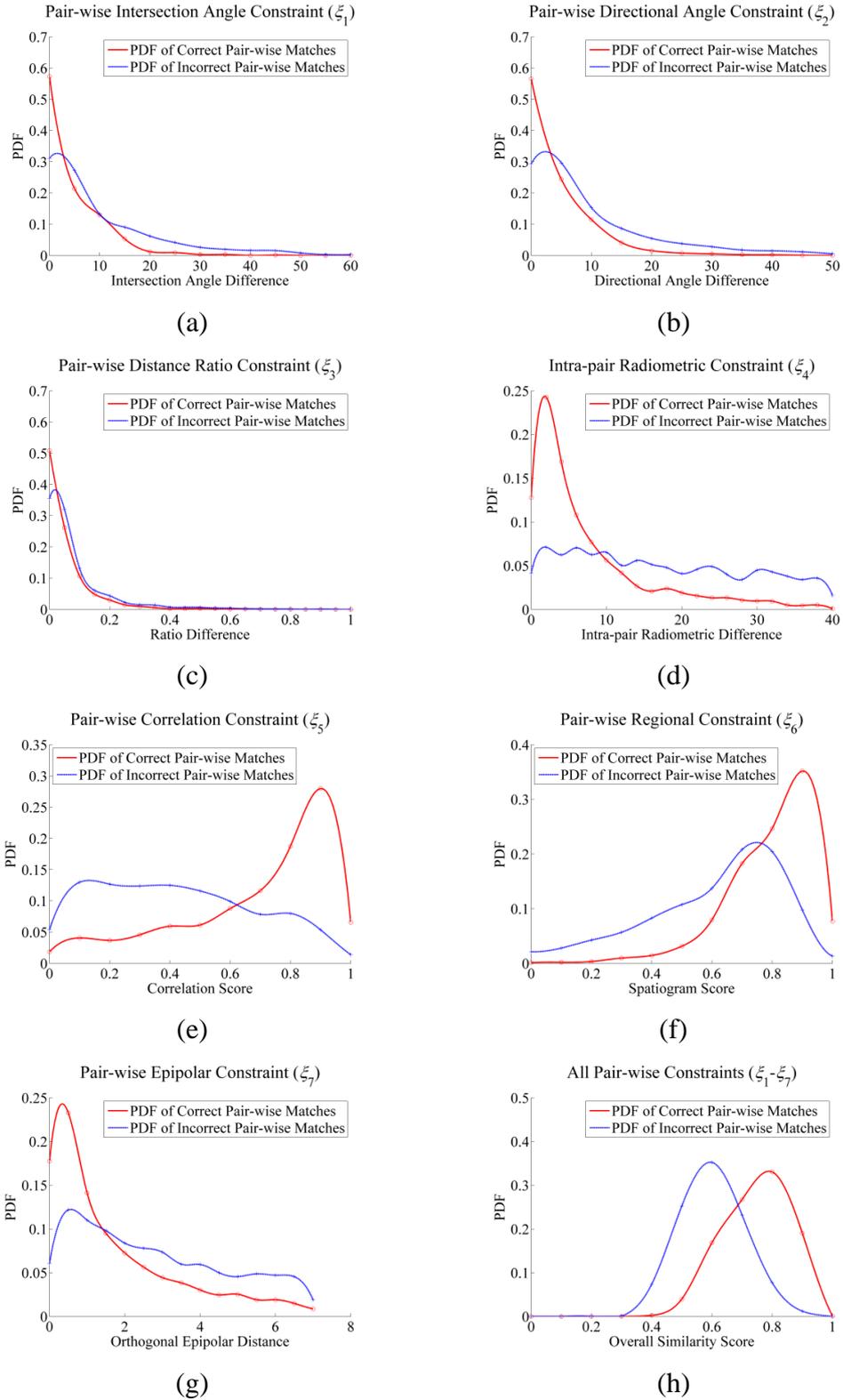


Figure 6.6 PDFs of pair-wise constraints and the final similarity function computed for the six Vaihingen test patches.

Table 6.1 Selected thresholds of the pair-wise constraints for the Vaihingen site.

<b>Pair-wise Constraints</b>	<b>Selected Threshold (<math>T_{\xi_1-\xi_7}</math>)</b>
Intersection Angle ( $\xi_1$ )	30°
Directional Angle ( $\xi_2$ )	30°
Distance Ratio ( $\xi_3$ )	0.4
Intra-pair Radiometric Value ( $\xi_4$ )	25
Correlation Score ( $\xi_5$ )	0.2
Regional Score ( $\xi_6$ )	0.4
Orthogonal Epipolar Distance ( $\xi_7$ )	5 pixels

remaining false matches with the selected geometric thresholds in Table 6.1. On the other hand, the PDF of the incorrect pair-wise matches revealed an interesting characteristic for the radiometric constraint ( $\xi_4$ ), being almost equally distributed between the radiometric extents (0-40). Thus, it is apparent that the selected radiometric threshold ( $T_{\xi_4} = 25$ ) can eliminate 30% of the remaining false pair-wise matches while providing most of the correct pair-wise matches. Similarly, the selected thresholds for the correlation and regional constraints may also have potential to eliminate approximately 30% and 20% of the false pair-wise matches, respectively. However, it is very important to state that most of those incorrect pair-wise matches may have a high possibility to be shared between different constraints. The PDF of overall similarity score indicates this fact (Fig. 6.6h). For all matches, a very nice bi-modal distribution with two distinct peaks is observed for the linear (equal-weighted) combination of the pair-wise constraints. The two normal distributions also have almost same variances with a mean similarity scores of 0.8 and 0.6 for the correct and incorrect pair-wise matches, respectively. However, it is clear that the overlapping part of the two normal distributions is so large that it can be easily represented by either class (correct or incorrect) mean value. Thus, as a result, the separation of the correct pair-wise matches from the incorrect ones is not further possible. However, the results given in Table 6.2 shows that those constraints have successfully reduced the total number of matching possibilities by  $\approx 80\%$ . Thus, the complexity of matching has also

Table 6.2 The total number of pair-wise relations observed before and after the initial pair-wise matching.

Processing Level	Number of Pair-wise Relations					
	Patch #1	Patch #2	Patch #3	Patch #4	Patch #5	Patch #6
Before Matching	2809	3627	1813	6032	3893	3194
After Initial Matching	478	637	484	1536	629	545
Reduction ratio	83%	82%	73%	75%	84%	83%

successfully reduced to an acceptable level.

### 6.3.1.2 Parameter Selection for the Precise Matching Stage

After setting up the thresholds provided in Table 6.1, the initial pair-wise matching stage is performed, and one-to-one line matches between stereo images are inferred from the entire pair-wise relations. After that, precise matching stage evaluates those one-to-one matches by means of three novel similarity measures, line-based Daisy ( $Sim_D$ ), Redundancy ( $Sim_R$ ), and Pair-wise Quality ( $Sim_Q$ ). At this point, since it is proved in the previous section that it is not possible to separate all correct pair-wise matches from the incorrect ones, we immediately investigate the PDF distributions of those one-to-one matches inferred from the pair-wise relations. Figure 6.7a-c illustrates the PDFs of the correct and incorrect one-to-one line matches for those three measures. It is apparent from the Fig. 6.7a that most of the correct line matches revealed Daisy similarity values ( $Sim_D$ ) of more than 0.8, whereas only a few portion of correct line matches show similarity values of less than 0.6. On the other hand, all of the incorrect line matches have a relatively constant PDF distribution among the similarity values. Thus, for instance, just a single Daisy similarity threshold of 0.8 will take out almost 80% of the incorrect matches after pair-wise matching while preserving 85% of the correct matches. A very similar decisive discrimination between the correct and incorrect line matches can also be seen from the PDFs computed for the

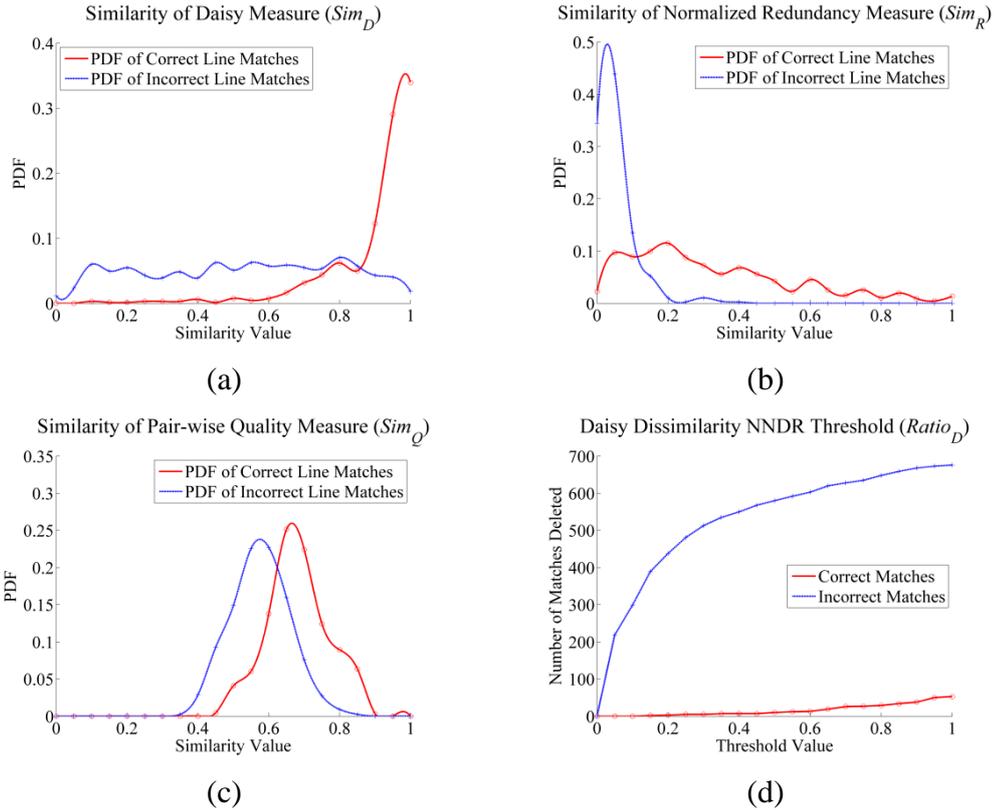


Figure 6.7 (a-c) PDFs of each measure utilized in the precise matching stage, and (d) the number of correct and incorrect matches deleted with respect to different Daisy dissimilarity NNDR thresholds.

normalized Redundancy measure ( $Sim_R$ ) (Fig. 6.7b). As strongly anticipated, the incorrect line matches have very low similarity values for the Redundancy measure; almost 95% of the incorrect matches have a normalized similarity value of less than 0.1, and almost all (99%) of the incorrect matches stay in a region defined by a maximal normalized Redundancy similarity value of 0.2. On the other hand, 80% of the correct line matches provided normalized Redundancy similarity value of larger than 0.1. Thus, a single similarity threshold 0.1 forced over the normalized Redundancy measure eliminates most of the incorrect line matches while keeping most of the correct matches. On the contrary to the Daisy and Redundancy distributions, the PDFs of the Pair-wise Quality measure ( $Sim_Q$ ) shown in Fig. 6.7c does not provide a good discrimination between the correct and incorrect matches. However, this is also an expected fact, since Pair-wise

Quality measure is computed by averaging the overall pair-wise similarity scores of each one-to-one line matches. Thus, it is not surprising to come up with a similar pattern for the overall pair-wise similarity scores illustrated in Fig. 6.6h. However, eventually, the quality measure also provides a hint about the correct and incorrect line matches.

One interesting fact for all those three PDFs figures is that it is not possible to eliminate a significant amount of incorrect line matches without removing any of the correct matches. For each case, a number of correct matches will be lost. Here, to avoid this, Daisy dissimilarity NNDR ( $1 - Sim_D$ ) metric in terms of the ratio of closest to second-closest matches of each line is investigated (Fig. 6.7d). The figure shows the number of correct and incorrect matches deleted based on various Daisy dissimilarity NNDR thresholds ( $ratio_D$ ). As can be seen from the Fig. 6.7d, a logarithmic increase of the number of incorrect matches proves that a huge number of incorrect line matches can be easily recognized by NNDR and can be immediately deleted with sacrificing only a relatively small number of correct matches. However, at this point, our aim is to delete only the line matches that are indisputably wrong; thus, we found that a strict dissimilarity NNDR threshold of 0.1 deletes almost 300 incorrect matches without removing any of the correct matches. As a result, a significant amount of false line matches can be deleted by Daisy dissimilarity NNDR before the iterative precise matching stage. Thereafter, we delete the line relations indicated by NNDR from the pair-wise matches, and for each match, we update the Redundancy and Pair-wise Quality measures. The PDFs of each measure after revising the Redundancy and Pair-wise Quality measures and the PDF of the overall similarity metric ( $Sim_T$ ) which is computed by the weighted combination  $\{w_D, w_R, w_Q\} = \{1/4, 1/2, 1/4\}$  of each measure is given in Fig. 6.8. If we visually compare the Fig. 6.7 with Fig.6.8, the differences between the PDFs of incorrect matches can be easily seen. Of course, the most important difference can be seen in Fig. 6.8c in which some of the incorrect matches have absolute zero value for their Pair-wise Quality metric. At this point, it should be pointed out that if a line relation in a pair is found to be

wrong (e.g. after NNDR), we do not directly delete the pair, since we don't have any inference (correct or incorrect) for the other match in that pair. However, if the other match in the pair is only observed in that wrong pair, since the Redundancy and Pair-wise Quality measures are computed in a pair-wise manner, that match will get zero values for the Redundancy and Pair-wise Quality measures. Thus, updating those pair-wise measures will always have a positive effect on the matching quality. We also continue to update Redundancy and Pair-wise Quality measures during the iterative processing; thus, the (updated) measures ( $Sim_R$  and  $Sim_Q$ ) turn out to be more reliable after each iteration.

Once the iterative matching procedure is completed, there is no ambiguity in the final matching list. However, the matching list may involve incorrect matches that may not violate any of the correct matches. To solve this problem, a final check with a global thresholding is performed with two global Daisy thresholds and

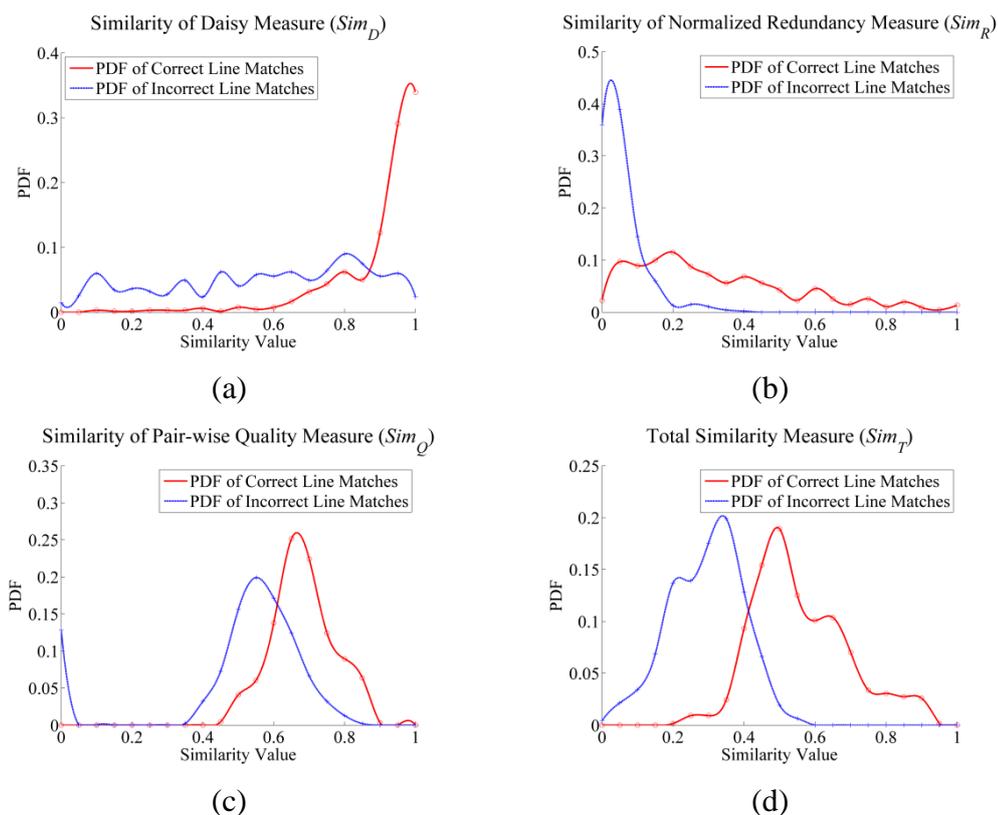


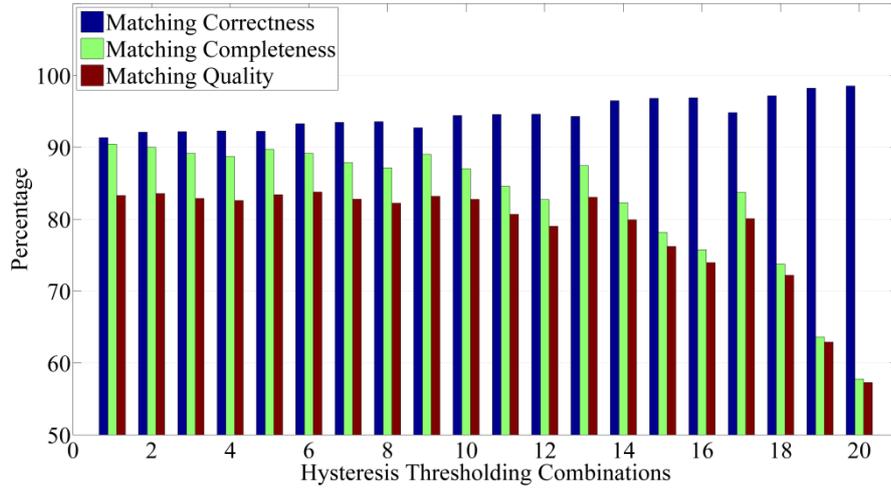
Figure 6.8 PDFs of each measure after enforcing the NNDR threshold of 0.1.

one Redundancy threshold (see Chapter 4, section 4.4.4):

- (i).  $Sim_D \geq Thr_D^1$
- (ii).  $Sim_D \geq Thr_D^2 \quad \& \quad Sim_R \geq Thr_R$

It is clear from the PDF of the correct matches, that a global Daisy threshold of  $Thr_D^1 \geq 0.1$  must be independently satisfied by every match. Thus, we eliminate all line-to-line matches that have Daisy similarity values of less than 0.1. For the second level of thresholding, increasing the Daisy threshold ( $Thr_D^2$ ) may have a high possibility to eliminate some of the correct matches as well, thus the elimination must also be controlled by a Redundancy threshold value of  $Thr_R$ . Thus, we tested a total of 20 different parameter combinations to analyze and understand the effect of the two thresholds utilized in the second level. We evaluated the final matching performances of each threshold combination in terms of the computed matching accuracy measures (matching correctness, completeness and quality). Fig. 6.9 presents the threshold values involved in the tests and the computed matching results for each combination. It is clear from the Fig. 6.9 that highest matching completeness values are achieved for the low Daisy threshold values. Actually, this can be easily validated by Fig. 6.8a, because only a small portion of the correct line matches have Daisy similarity values of less than 0.55. However, it is also clear that at the same level, the correctness values have the lowest matching percentages; thus, it states that the incorrect matches could not be eliminated efficiently. On the other hand, increasing the Daisy threshold to higher values yield better correctness results; however, at the same time, the completeness levels reached depends on the Redundancy thresholds utilized. As can be seen easily from the Fig. 6.9, high Daisy thresholds combined with high Redundancy thresholds cause a significant increase in terms of the matching correctness (up to 99%), whereas a significant decrease in terms of the matching completeness (down to 58%) is inevitable. Indeed, matching quality values reveals a good balance between the values of the matching correctness and completeness. It is obvious that the highest quality values computed are on the

Hysteresis Parameters vs Accuracy ( $Thr_D^2, Thr_R$ )



Comb ID	$Thr_D^2$	$Thr_R$	Comb ID	$Thr_D^2$	$Thr_R$
#1	0.55	0.1	#11	0.75	0.3
#2	0.55	0.2	#12	0.75	0.4
#3	0.55	0.3	#13	0.85	0.1
#4	0.55	0.4	#14	0.85	0.2
#5	0.65	0.1	#15	0.85	0.3
#6	0.65	0.2	#16	0.85	0.4
#7	0.65	0.3	#17	0.95	0.1
#8	0.65	0.4	#18	0.95	0.2
#9	0.75	0.1	#19	0.95	0.3
#10	0.75	0.2	#20	0.95	0.4

Figure 6.9 The results of the performance test based on various hysteresis thresholding combinations for the Vaihingen test site.

order of 83%. For line matching, our aim is to first minimize the number false matches (thus to increase the correctness level) while preserving the highest level of completeness. Therefore, according to the results, that sense of balance is provided by the threshold combination #13 in which the Daisy and Redundancy thresholds are defined as 0.85 and 0.1, respectively. Actually, if we analyze the Fig. 6.8a and b, those threshold values can also be interpreted from those figures. It is evident that both thresholds are very close to the intersection points of the

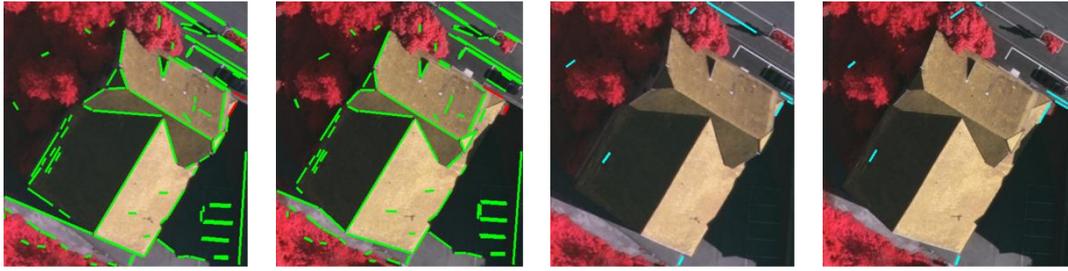
Table 6.3 The selected thresholds for the precise matching of the Vaihingen site.

Precise Matching Parameters	Selected Threshold
Daisy dissimilarity NNDR ( $ratio_D$ )	0.1
Global Daisy Threshold – 1 ( $Thr_D^1$ )	0.1
Global Daisy Threshold – 2 ( $Thr_D^2$ )	0.85
Global Redundancy Threshold ( $Thr_R$ )	0.1

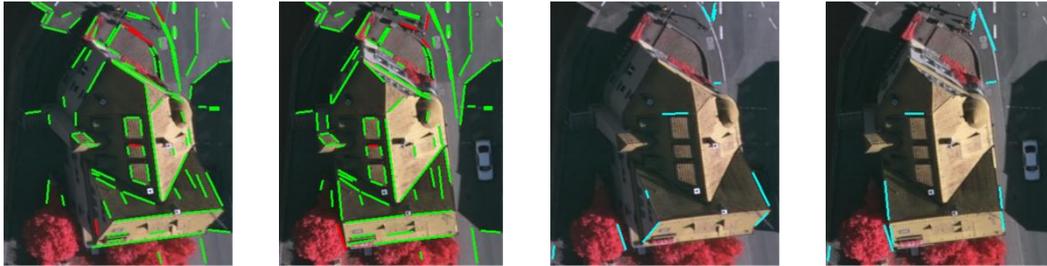
PDFs of the correct and incorrect line matches. Accordingly, for the Vaihingen test site, we fix all the required parameters for the precise matching stage provided in Table 6.3.

### 6.3.1.3 Evaluation and Discussion of the Line Matching Results

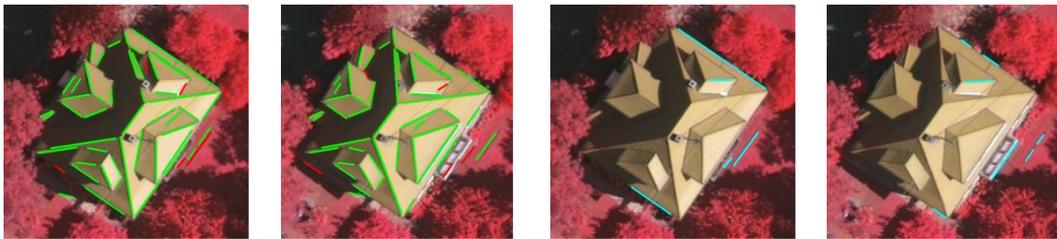
To evaluate the matching performance of the proposed line matching approach, in addition to the 6 stereo test patches shown in Fig. 6.5, we selected 9 more test patches from the Vaihingen test site (Fig. 6.10). All those 15 image patches are selected based on their various built-up characteristics (dense-sparse environments, flat-gable-complex roofs styles, repetitive linear patterns, constant-irregular height variations, moving objects etc.). During the experiments, for all test patches, we applied a 50 m ( $\approx 162$  pixels for the Vaihingen images) search range difference along the epipolar lines. In order to perform a detailed matching analysis, for all test patches, a line-to-line reference matching list is generated by visual inspection. Thus, for each test site, the number of correct and incorrect line matches is computed automatically by comparing the automated matching results with the manually generated reference matching list. The numerical and visual results of the correct and incorrect line matches are provided in Fig. 6.10. In order to further assess the quality level of the line matching, we further analyzed the number of line matches missed by the approach; thus, the final line matching results in terms of the correctness, completeness and quality levels for each test patch are also provided (Fig. 6.11). The numerical and visual results of the missed line matches can also be found in Fig. 6.10.



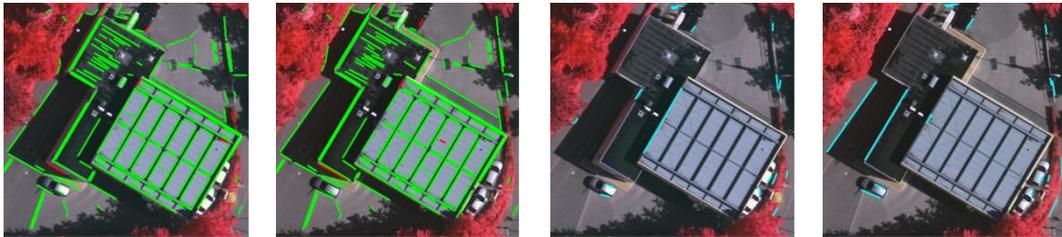
Test Patch #1 - Matched Lines: 101, Incorrect: 2, Missed: 8



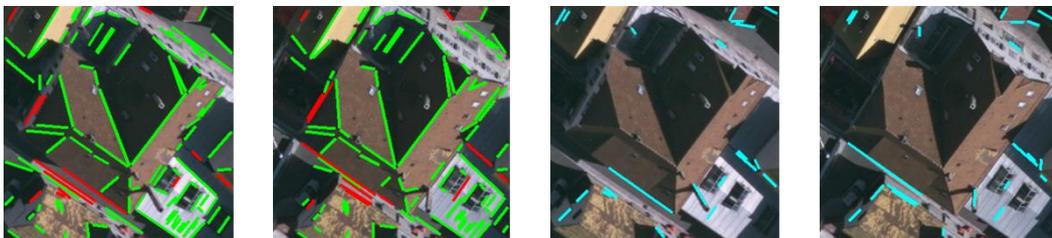
Test Patch #2 - Matched Lines: 95, Incorrect: 5, Missed: 11



Test Patch #3 - Matched Lines: 55, Incorrect: 6, Missed: 7



Test Patch #4 - Matched Lines: 165, Incorrect: 5, Missed: 16

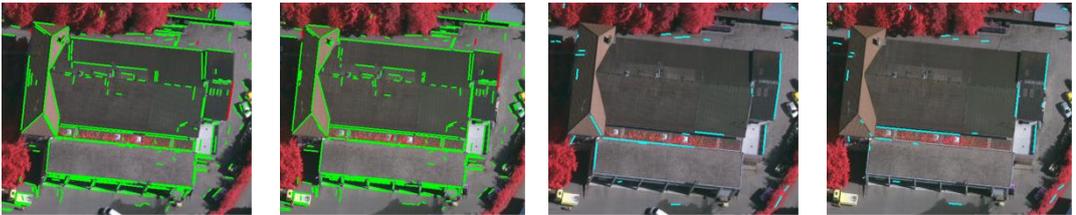


Test Patch #5 - Matched Lines: 113, Incorrect: 13, Missed: 21

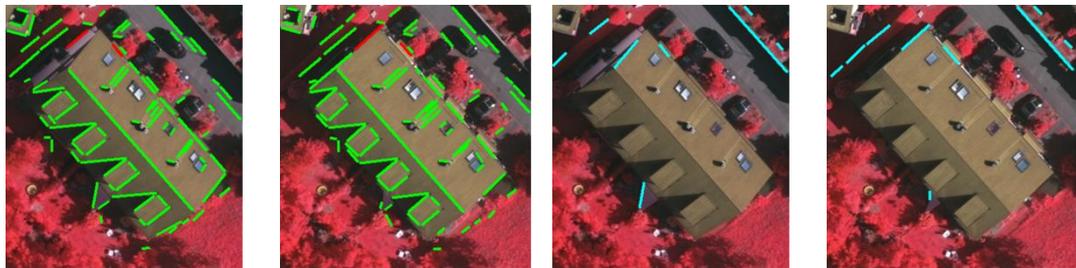
Figure 6.10 The results of the proposed approach. The colors green, red, and cyan correspond to correct, incorrect and missed matches, respectively.



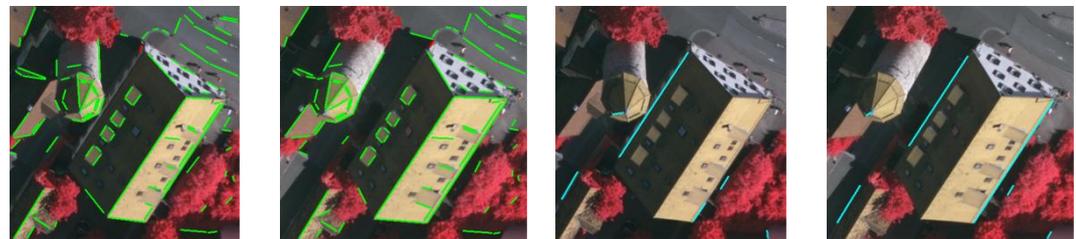
Test Patch #6 - Matched Lines: 125, Incorrect: 2, Missed: 27



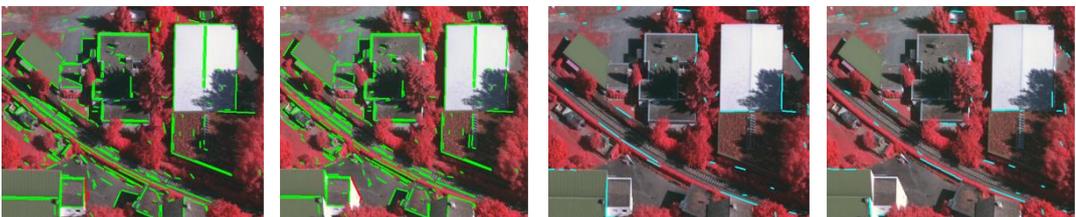
Test Patch #7 - Matched Lines: 255, Incorrect: 4, Missed: 30



Test Patch #8 - Matched Lines: 87, Incorrect: 2, Missed: 9

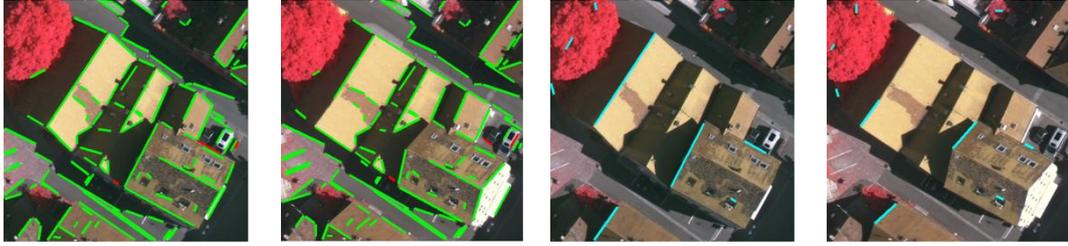


Test Patch #9 - Matched Lines: 101, Incorrect: 2, Missed: 6

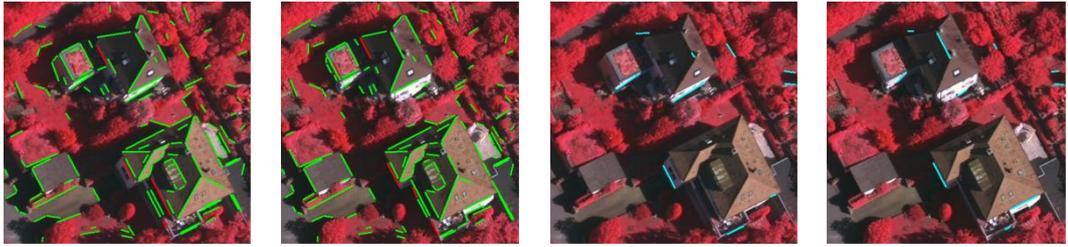


Test Patch #10 - Matched Lines: 270, Incorrect: 3, Missed: 40

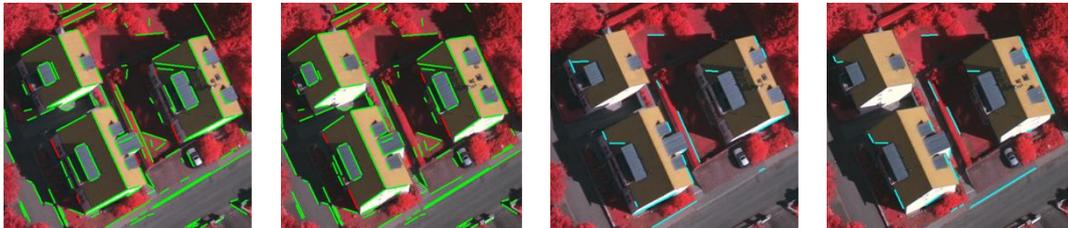
Figure 6.10 (continued)



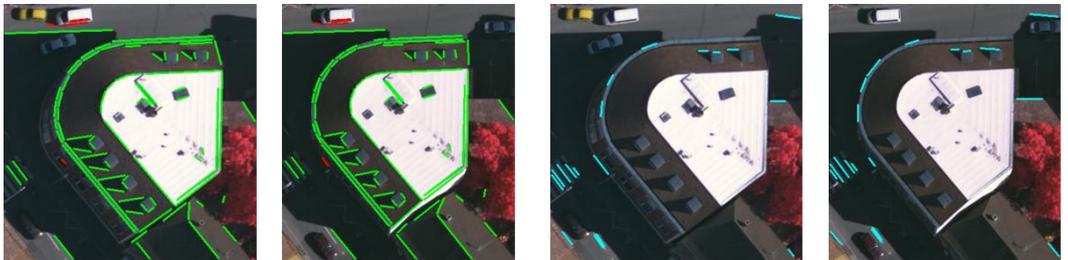
Test Patch #11 - Matched Lines: 115, Incorrect: 3, Missed: 9



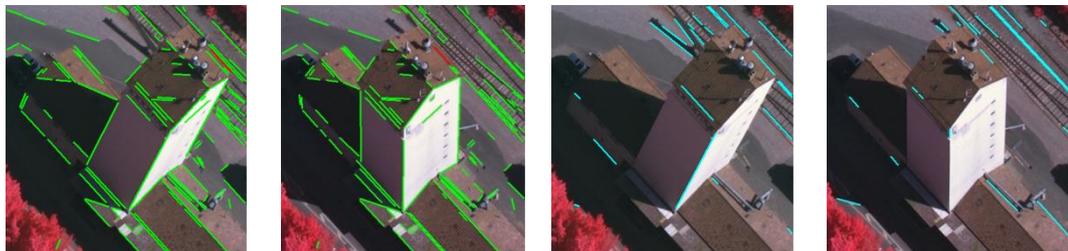
Test Patch #12 - Matched Lines: 140, Incorrect: 6, Missed: 10



Test Patch #13 - Matched Lines: 142, Incorrect: 8, Missed: 21

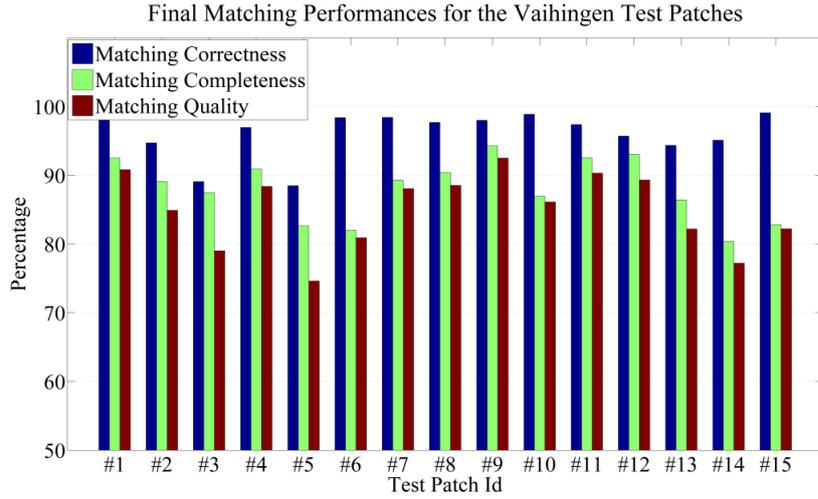


Test Patch #14 - Matched Lines: 82, Incorrect: 4, Missed: 19



Test Patch #15 - Matched Lines: 112, Incorrect: 1, Missed: 23

Figure 6.10 (continued)



Patch Id	Correctness (%)	Completeness (%)	Quality (%)
	$\left[ \frac{TP}{TP + FP} \right]$	$\left[ \frac{TP}{TP + FN} \right]$	$\left[ \frac{TP}{TP + FP + FN} \right]$
#1	98.0	92.5	90.8
#2	94.7	89.1	84.9
#3	89.1	87.5	79.0
#4	96.9	90.9	88.4
#5	88.5	82.6	74.6
#6	98.4	82.0	80.9
#7	98.4	89.3	88.1
#8	97.7	90.4	88.5
#9	98.0	94.2	92.5
#10	98.9	86.9	86.1
#11	97.4	92.6	90.3
#12	95.7	93.1	89.3
#13	94.4	86.5	82.2
#14	95.1	80.4	77.2
#15	99.1	82.8	82.2
<b>Patches 1-6 (<math>\mu \pm \sigma</math>)</b>	<b>94.3 <math>\pm</math> 4.4</b>	<b>87.5 <math>\pm</math> 4.3</b>	<b>83.1 <math>\pm</math> 6.1</b>
<b>Patches 7-15 (<math>\mu \pm \sigma</math>)</b>	<b>97.2 <math>\pm</math> 1.7</b>	<b>88.5 <math>\pm</math> 4.7</b>	<b>86.3 <math>\pm</math> 4.9</b>
<b>Overall (<math>\mu \pm \sigma</math>)</b>	<b>96.0 <math>\pm</math> 3.3</b>	<b>88.1 <math>\pm</math> 4.4</b>	<b>85.0 <math>\pm</math> 5.4</b>

Figure 6.11 The final matching performances of the proposed approach computed for the 15 test patches selected from the Vaihingen test site.

According to the matching results presented in Fig. 6.11, the proposed approach revealed very high rates of correctness values ranging between 88% and 99%. Among all 15 test patches, 13 of them have reached correctness levels of over 94%, thus the overall correctness performance of matching is computed to be 96% with a standard deviation of 3.3%. If the complexities of the test sites are taken into account, this seems to be a very good matching performance. If we visually examine the incorrect line matches in detail, most of them appear in the occluded parts that are only visible in one of the stereo images. This is due to the reason that if a line segment in one of the images has no corresponding line segment in the other image (especially the case of occluded parts), that line segment can still be matched with a line segment in the other image. Although we apply a two level global thresholding to reduce the number of incorrect matches that may occur due to this reason; unfortunately, it is not always possible to fully remove the erroneous matches in a stereo environment. One other reason to cause incorrectly matched line segments is moving objects such as cars, pedestrians etc. in the images. Although there is only a very short time difference during the acquisitions between the in-strip stereo images, the moving objects may also cause incorrect line matches. Various examples of moving cars are visible in different parts of image patches #2, #7 and #14, and walking of several pedestrians is recognizable in patches #6, #7 and #9. However, among all those image patches, the erroneous line matches due to moving objects is only observed in patch #14 (the white vehicle in the top-left corner) and no incorrect matches are identified in the other concerned patches (#2, #6, #7, and #9). This is related with the directional movement of the white vehicle in patch #14 which is in that case from right to left. Thus, this directional movement coincides with the flight path which is also the same direction of the epipolar search region utilized during the matching stage. Therefore, although the three line segments matched correspond to exactly the correct parts of the white vehicle in patch #14; since the object is moving, the matching line segments are no longer representing the correct disparities, and they are labeled as incorrect.

The completeness values of the line matches for the test patches are also revealed remarkable results ranging between 80% and 94%. The overall completeness value is computed to be 88.1% with a standard deviation of 4.4%. Similar to the level of correctness computed, it is believed that the completeness levels reached in each patch is also relatively good. On the other hand, a number of correct line matches are missed for each stereo image patch. Therefore, at this point, we believe that it is essential to examine missed matches in detail to fully understand the reasons behind those missed matches. Therefore, we carefully investigated all 15 patches and as a result, the possible reasons can be categorized into six different groups:

- i.* line segments that cannot be paired with any other line segments,
- ii.* line segments that are exactly aligned with the epipolar line,
- iii.* line segments whose left and right flanking regions are both occluded from other objects due to perspective distortion,
- iv.* line segments that correspond to multiple non-collinear line segments in one of the stereo images,
- v.* line segments that correspond to multiple very short collinear line segments in one of the stereo images,
- vi.* line segments that are very short,

It is evident from most of the test patches that a large portion of the missed matches belongs to the first category. In general, this is probably the reason most expected, since the line segments that are located in the corners and the sides of the images have relatively less opportunity to get paired with other nearby line segments. For sure, this type of missed matches can be reduced by increasing the proximity distance ( $T_{\text{prox}}$ ) utilized during the pair generation step; however, as we already demonstrated in section 6.3.3.1, increasing the proximity distance will result in an enormous number of line pairs which significantly increases the matching complexity. For example, in the stereo patch #8, 4 out of 6 missed matches (three line matches in the top-right corner and one line match in the top-

left corner) occur due to this reason. Although those line matches are very distinct and clear, it is impossible to match those line segments with a pair-wise logic, because there are not any line segments in their close neighborhoods to form a line pair. Thus, this is one of the major limitations of the pair-wise matching logic, and we can conclude that in a pair-wise approach, the lines that do not have any consistent line segments in their neighborhoods cannot be matched.

The second category of missed matches involves the line segments that are exactly aligned with the epipolar line. Currently, the proposed line matching approach requires only a very slight difference ( $\theta_{\text{diff}} > 0^\circ$ ) between the line segments and the epipolar line. However, if the line segments are exactly aligned with the epipolar line ( $\theta_{\text{diff}} = 0^\circ$ ), the quadrilateral region that is utilized to collect matching candidates turns into a line segment; thus, the correct matching candidates on the other image may not be collected properly. Moreover, it is impossible to perform a point-to-point correspondence for those cases, and therefore, the reliability of some of the pair-wise constraints decreases dramatically. Some examples of the missed matches due to this reason are noticed in stereo patch #14. Since the epipolar direction is from right to left, four line segments that are exactly aligned with the epipolar line could not be matched.

One interesting reason for losing some of the correct matches is due to their dissimilarity of the left and right flanking regions at the same time caused by different perspective views. In aerial images, in most of the cases, at least one side of the corresponding line segments in stereo aerial images is similar. However, in exceptional cases, this rule can be violated. The long line segment in the central part of the stereo patch #9 is a very good example for this type of violation. For that case, in the left image, one neighborhood of the line segment is occluded by the tower located in the vicinity of the building. On the other hand, in the right image, the other neighborhood of the line segment is occluded by the shed dormers of the building. As a result, both sides of the line segments in stereo

images are different from each other, and the rejection of the match is unavoidable.

The fourth category of missed matches occurs due to the accidental merging of the line segments during the stage of straight line extraction. Typically, the line segments observed for the buildings that have gable roof structures belong to this group. The example given in the gable roof located in the bottom-right corner of the stereo patch #11 clarifies the problem. In the left view of stereo patch #11, the two line segments belonging to the two sides of the gable roof are accidentally merged into a single line segment due to perspective view. On the other hand, in the right view of the patch #11, they are successfully found as two separate line segments. In the proposed line matching approach, although the initial pair-wise matching stage can successfully match an individual line segment extracted in one view to the multiple non-collinear segments in the other view, the precise matching stage only allows the matching formations for the multiple line segments consisting of only collinear lines. For that specific case, apparently, the collinearity of the two matching segment is violated; thus, as can be seen in the results presented in Fig. 6.10 - patch #11, only one of those matches which has the highest matching similarity score is matched successfully, and the other match is deleted from the matching list. A different example can also be seen for the gable roof shown in the left-bottom corner of the same stereo patch which also suffers from the same problem.

In contrast to the fourth category, the fifth category of missed matches occurs due to the accidental fragmentation of the line segments during the stage of straight line extraction. Indeed, in this category, we encounter two different reasons which are responsible for the missed matches. The two line matches missed in the left-center part of the stereo patch #4 illustrate clear examples for this category. The first problem is related with the pair-wise logic, some of the fragmented short line segments may not be paired successfully with other line segments. Thus, the matching of those line segments is not possible. The multiple segments that

belong to the shadow of the building in the stereo patch #4 are good examples for this case. As can be seen in the figure, the three line segments in the left view correspond to a single segment in the right view. However, the short segment in the left view is not matched with the segment in the right-view, since it cannot be paired with any other line segment nearby and failed to match. The second problem of this category is related with the difficulties that may occur during the verification of the collinearity of two or more line segments in image space. The example in the stereo patch #4 clarifies this problem. In the left-view, the boundary of the building is split into two line segments, whereas the corresponding line segment in the right view is found as a single segment. In that case, it is clear that there is a slight angle difference between the short and the long segments shown in the left view, thus those two lines cannot be verified as a collinear line segments in image space. As a result, the correct match with the short segment in the left view is missing.

The sixth category is related with the line segments that are very short. During the line matching stage, we allow very short line segments (a minimum length threshold of 10 pixels is utilized for all cases) to capture the highest level of detail ( $\approx 80$  cm for the test site Vaihingen) in the aerial images. However, the uncertainty of the short line segments, especially in the orientation component may have a major negative effect during the computation of some of the pair-wise constraints especially the epipolar constraint. Since the orthogonal distances are evaluated during the epipolar constraint, any orientation error may cause considerable positional errors during the computation of the point of intersection of a pair; hence computed orthogonal distances may also contain substantial errors. As a result, the correct matching candidate pair in the other image may not be correctly identified and may be lost. One other problem is that, some of the short line segments are observed in the cases in which they belong to a natural object such as vegetation. As can be seen in stereo patch #1, although most of the correct line matches on the vegetation objects are successfully identified, it is not always possible to recover all of those correct line matches detected over a

vegetation object. This is certainly related with the pair-wise logic: the line segments extracted on the vegetation objects must be correctly paired in order to be accurately matched with their corresponding line segments.

In overall, for the proposed approach, a line matching quality level of approximately 85% is achieved for the Vaihingen test patches (Fig 6.11). It is believed that, if the complexities of the test patches are taken into account, this seems to be a very good performance. Furthermore, on the contrary to most of the previous line matching approaches, we do not impose any external dataset to the matching (stereo DSM, additional views etc.) to solve the line matching ambiguities. It should also be emphasized that the curved segments (belonging to buildings, roads, etc.) that can be piece-wise linear approximated are also matched successfully. The results of the stereo test patch #14 prove this issue. Actually, this is not a surprising fact, since the piece-wise approximated linear segments are also particularly suitable to be matched by the proposed pair-wise approach.

### **6.3.2 The Results and Discussion of the Hannover Test Site**

#### **6.3.2.1 Parameter Selection for the Initial Pair-wise Matching Stage**

In this part, to determine the required parameters for the pair-wise matching of the Hannover test site, the same strategy presented for the Vaihingen test site is followed. Six test patches selected from the different parts of the Hannover test site are used to analyze and understand the behavior of the pair-wise parameters (Fig. 6.12). Similar to the Vaihingen test site, the proximity distance ( $T_{prox}$ ) is fixed to 8 m on object space for the pair-wise matching of all test patches. During the assessments, PDF for the correct pair-wise matches and as well as incorrect pair-wise matches are computed for the selected six test patches. Fig. 6.13 shows the computed PDFs of the correct and incorrect pair-wise matches that are generated using the six image pairs given in Fig. 6.12. As can be seen from Fig. 6.13, the PDFs of the correct and incorrect pair-wise matches are quite similar to

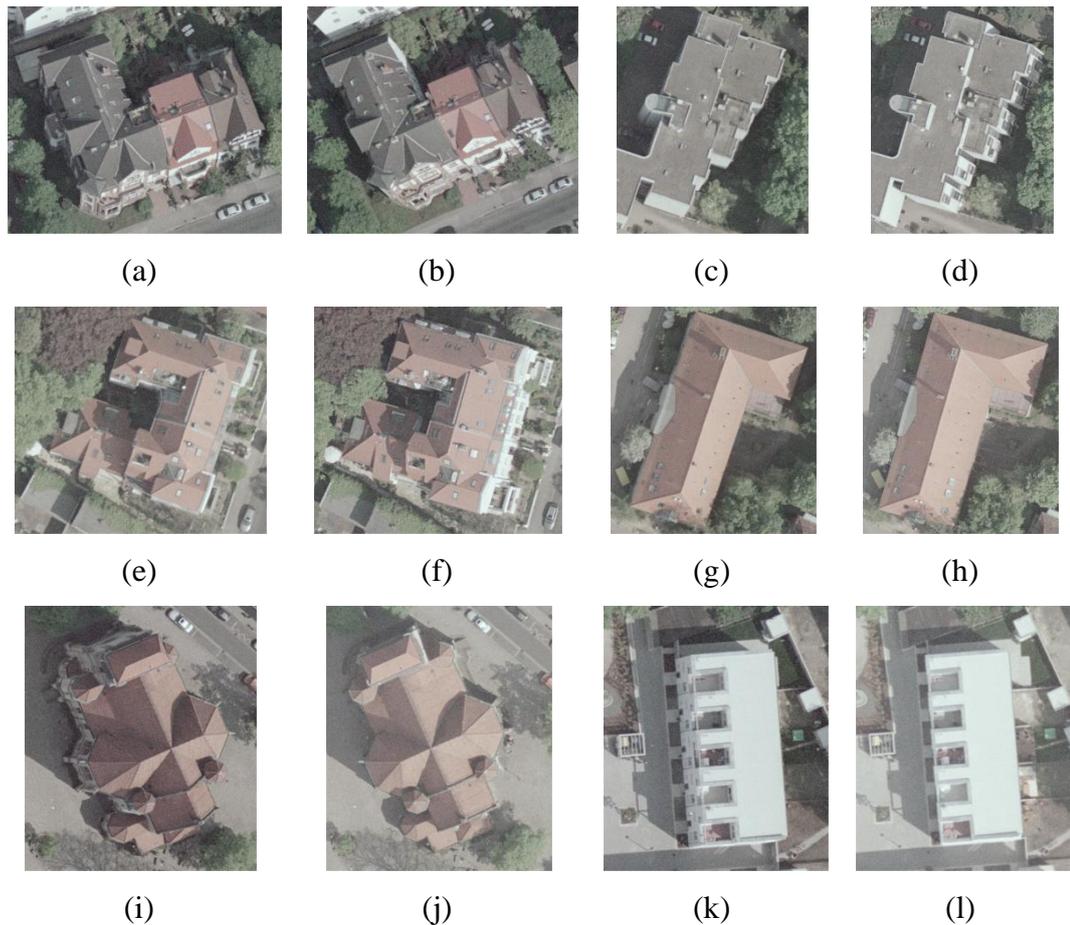


Figure 6.12 Six stereo test patches selected from the Hannover test site.

the PDF distributions of the Vaihingen test patches shown in Fig. 6.6. Actually, this is not a surprising fact due to the similar image acquisition geometry. Although the average image baseline distance of the Hannover test site is two times larger than the average baseline distance of the Vaihingen test site, the stereo images of the two test sites are acquired with very similar B/H ratios. Thus, similar PDF distributions are strongly anticipated. However, the radiometric quality differences of the images (digital vs. analogue) of the two test sites caused an interesting effect on the PDFs computed for the correlation constraint ( $\zeta_5$ ). It is clear from the Fig. 6.13e that the correlation scores of the correct pair-wise matches are negatively affected. On the other hand, the scores of the incorrect pair-wise matches are clearly reduced to the desired level which actually is not observed for the correlation scores of the incorrect pair-wise matches computed

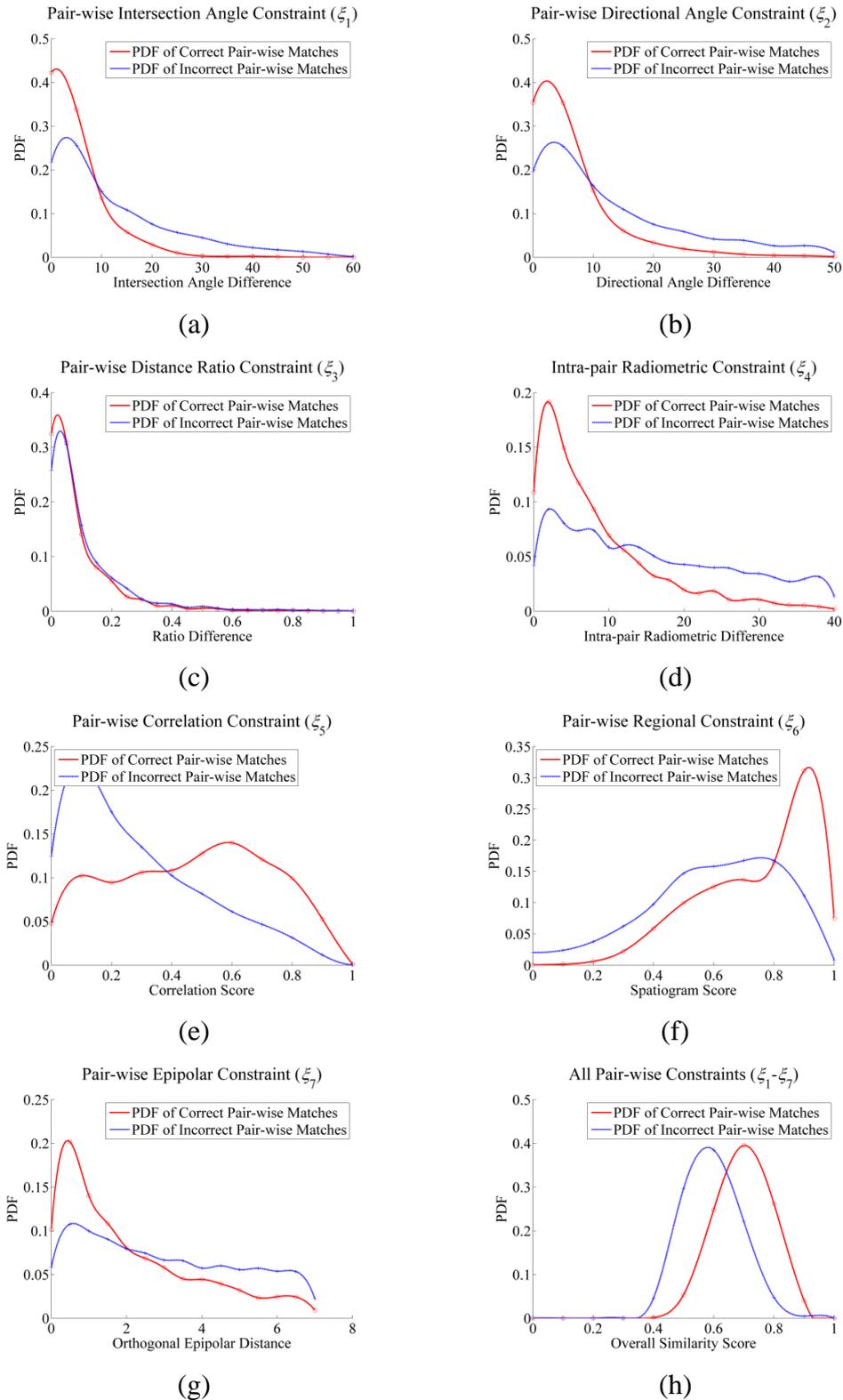


Figure 6.13 PDFs of each pair-wise constraint and the final similarity function computed for the six Hannover test patches.

for the Vaihingen test site (Fig 6.6e). Nevertheless, a correlation threshold of ( $T_{\xi_4} = 0.2$ ) still seems to be a good choice. Thus, for the Hannover test site, we also utilize the same thresholds utilized for the Vaihingen test site (in Table 6.1), since those thresholds seem to be valid and appropriate for the pair-wise matching of Hannover images as well.

### 6.3.2.2 Parameter Selection for the Precise Matching Stage

We also applied the similar strategy to evaluate the PDF distributions of those one-to-one matches inferred from the pair-wise relations. Figure 6.14a-c illustrate the PDFs of the correct and incorrect one-to-one line matches for the three similarity measures (line-based Daisy ( $Sim_D$ ), Redundancy ( $Sim_R$ ), and Pair-wise

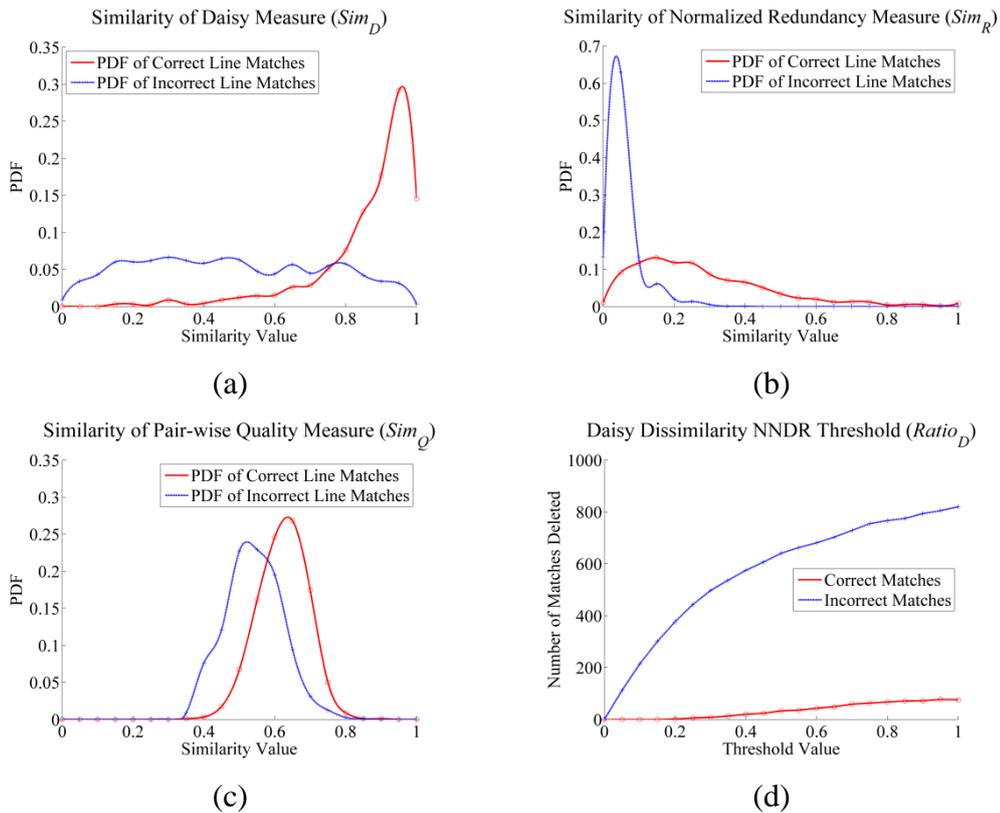


Figure 6.14 (a-c) PDFs of each measure utilized in the precise matching stage, and (d) the number of correct and incorrect matches deleted with respect to different Daisy dissimilarity NNDR thresholds.

Quality ( $Sim_Q$ ) computed for the six Hannover test patches. Once again, the PDF results of the Hannover test patches bear a clear resemblance to the PDF results obtained from Vaihingen test patches. The PDFs of correct and incorrect matches in Fig. 6.14a, b reveals that the measures of line-based Daisy ( $Sim_D$ ) and the normalized Redundancy ( $Sim_R$ ) are robust to noise inherent in the analogue images. The results in Fig. 6.14d proves that a huge number of incorrect line matches can still be easily recognized by NNDR. However, at this point, on the contrary to the Vaihingen test site results, a dissimilarity NNDR threshold ( $ratio_D$ ) of 0.15 seems to be the optimum threshold value to eliminate the incorrect matches without removing any of the correct matches. However, the strict threshold value ( $ratio_D = 0.1$ ) selected for the Vaihingen site also achieves to eliminate more than 200 indisputably wrong line-to-line relations; thus, for the sake of simplicity, we also continue to utilize the same Daisy dissimilarity NNDR threshold of 0.1 for the Hannover test site as well. After removing the line

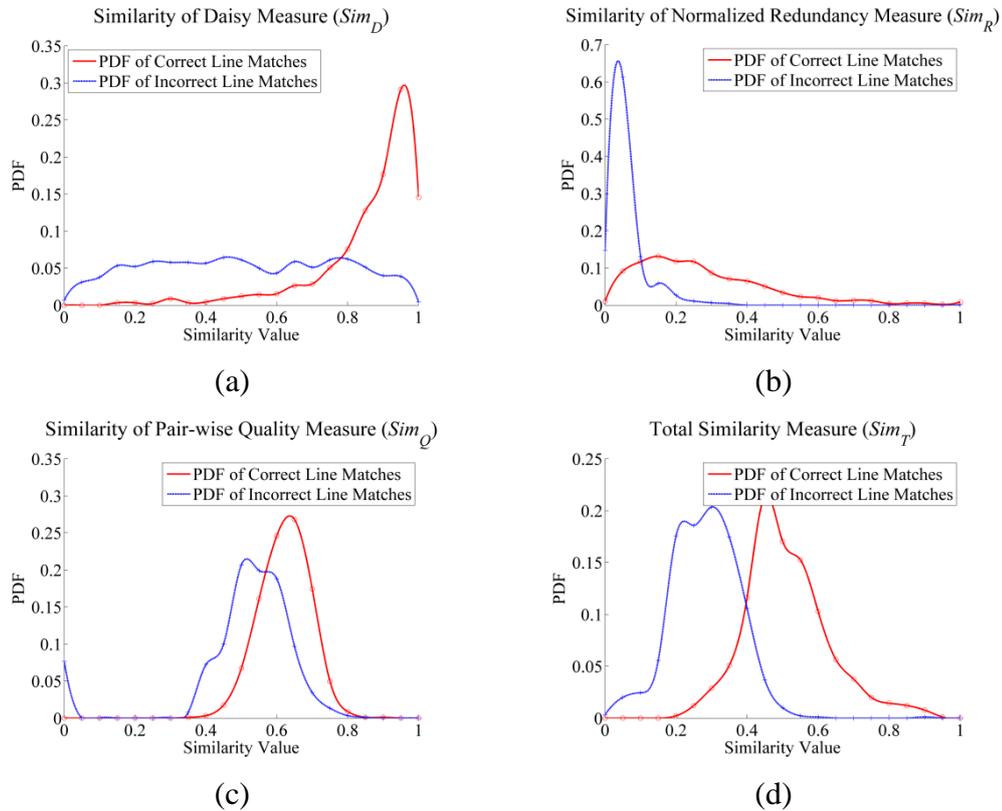


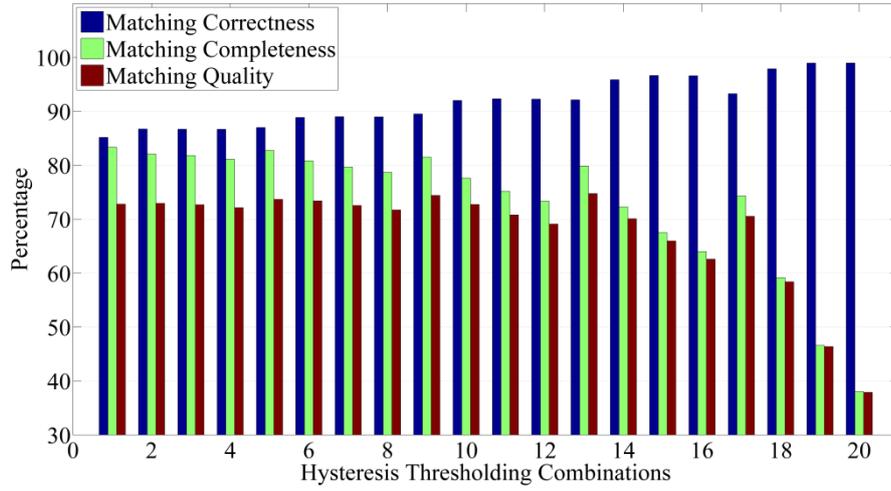
Figure 6.15 PDFs of each measure after enforcing the NNDR threshold of 0.1.

relations indicated by NNDR from the pair-wise matches, and successively updating the Redundancy and Pair-wise Quality measures, the resulting PDFs of each measure and the PDF of the overall similarity metric ( $Sim_T$ ) which is computed by the weighted combination  $\{w_D, w_R, w_Q\} = \{1/4, 1/2, 1/4\}$  of each measure is given in Fig. 6.15. If we visually compare the results in Fig. 6.15 with results in Fig. 6.8, the similar trend between the Hannover and Vaihingen datasets are also clearly visible and evident. For the hysteresis global thresholding, once again, we investigated a total of 20 different parameter combinations and evaluated the final matching performances of each threshold combination in terms of the computed matching accuracy measures (matching correctness, completeness and quality). Fig. 6.16 presents the threshold values involved in the tests and the computed matching results for each combination. As the results indicated, the hysteresis thresholds selected for the Vaihingen test site is also found to be the best combination for the Hannover test site; thus, for the precise matching of Hannover test patches, we utilized the same thresholds provided in Table 6.3.

### 6.3.2.3 Evaluation and Discussion of the Line Matching Results

To evaluate the matching performances of the proposed line matching approach, in addition to the 6 stereo test patches shown in Fig. 6.12, we selected 9 more test patches from the Hannover test site (Fig. 6.17). Similar to the previous experiments, for all test patches, we applied a 50 m ( $\approx 216$  pixels for Hannover images) search range difference along the epipolar lines and all the results are evaluated with respect to a line-to-line reference matching list generated by visual inspection. The numerical and visual results of the correct, incorrect and missed line matches are provided in Fig. 6.17. The line matching results of the Hannover test patches in terms of the correctness, completeness and quality levels are also provided in Fig. 6.18.

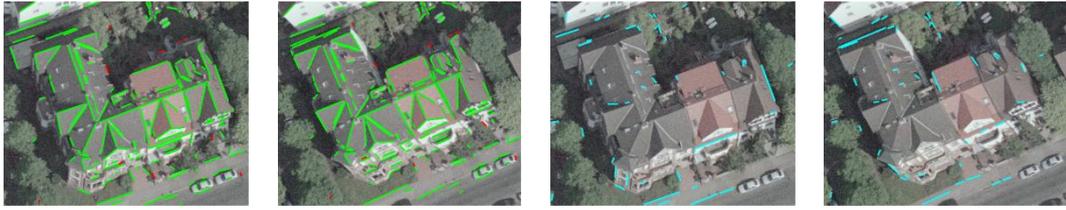
Hysteresis Parameters vs Accuracy ( $Thr_D^2, Thr_R$ )



Comb ID	$Thr_D^2$	$Thr_R$	Comb ID	$Thr_D^2$	$Thr_R$
#1	0.55	0.1	#11	0.75	0.3
#2	0.55	0.2	#12	0.75	0.4
#3	0.55	0.3	#13	0.85	0.1
#4	0.55	0.4	#14	0.85	0.2
#5	0.65	0.1	#15	0.85	0.3
#6	0.65	0.2	#16	0.85	0.4
#7	0.65	0.3	#17	0.95	0.1
#8	0.65	0.4	#18	0.95	0.2
#9	0.75	0.1	#19	0.95	0.3
#10	0.75	0.2	#20	0.95	0.4

Figure 6.16 The results of the performance test based on various hysteresis thresholding combinations for the Hannover test site.

According to the matching results presented in Fig. 6.18, similar to the results of the Vaihingen test patches, the proposed approach revealed very high rates of correctness values ranging between 90% and 98%. The overall correctness performance of the matching is computed to be 94.5% with a standard deviation of 2.5%. If the complexities of the test sites and the noise level of the images are taken into account, we believe that this is an impressive stereo line matching performance. On the other hand, the completeness values of the line matches for



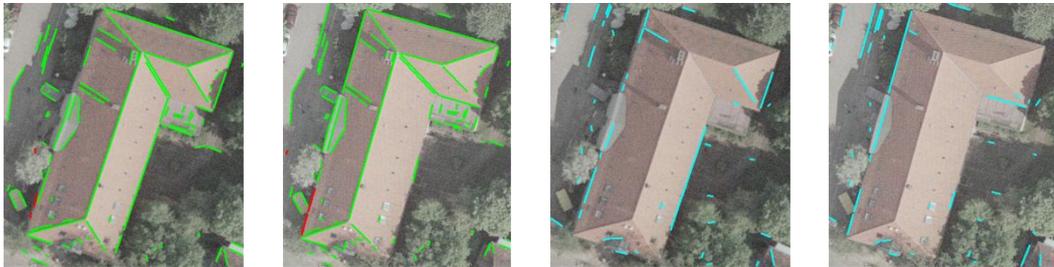
Test Patch #1 - Matched Lines: 185, Incorrect: 10, Missed: 44



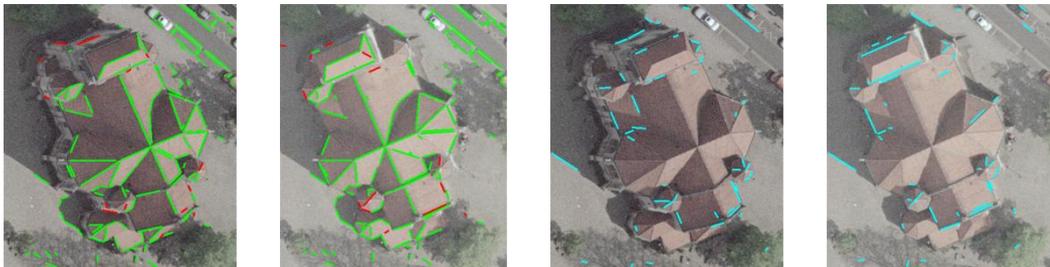
Test Patch #2 - Matched Lines: 110, Incorrect: 9, Missed: 28



Test Patch #3 - Matched Lines: 186, Incorrect: 18, Missed: 45

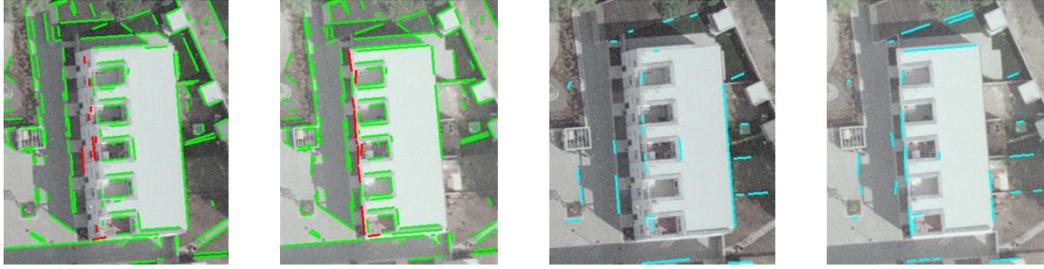


Test Patch #4 - Matched Lines: 99, Incorrect: 3, Missed: 42



Test Patch #5 - Matched Lines: 154, Incorrect: 14, Missed: 47

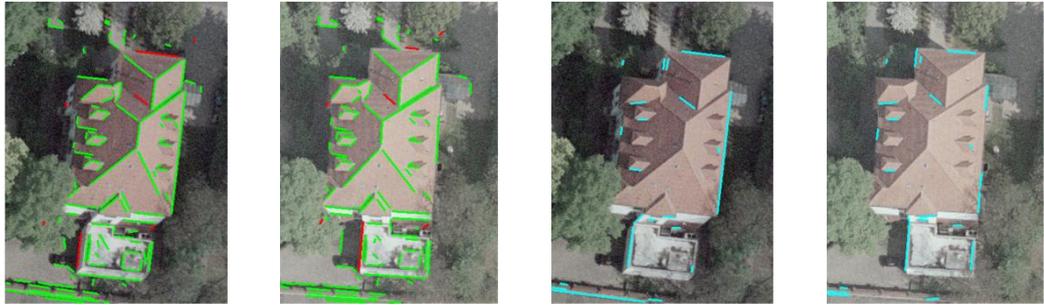
Figure 6.17 The results of the proposed approach. The colors green, red, and cyan correspond to correct, incorrect and missed matches, respectively.



Test Patch #6 - Matched Lines: 157, Incorrect: 10, Missed: 31



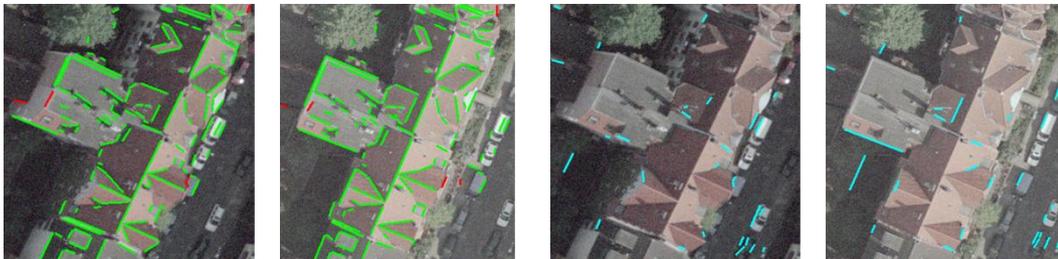
Test Patch #7 - Matched Lines: 166, Incorrect: 6, Missed: 39



Test Patch #8 - Matched Lines: 120, Incorrect: 7, Missed: 24

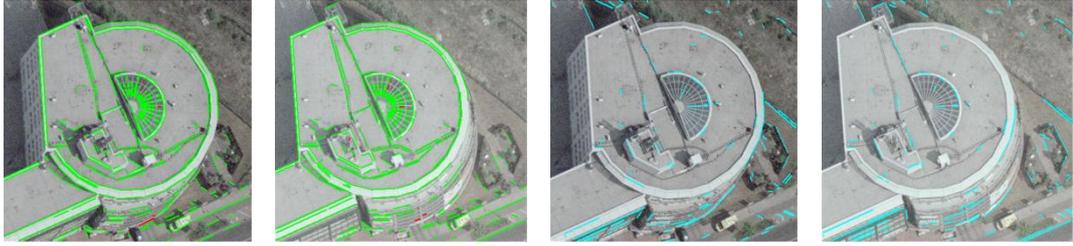


Test Patch #9 - Matched Lines: 188, Incorrect: 9, Missed: 43

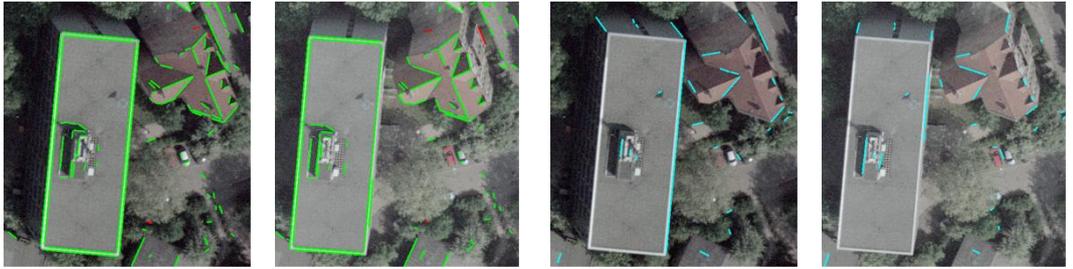


Test Patch #10 - Matched Lines: 142, Incorrect: 5, Missed: 28

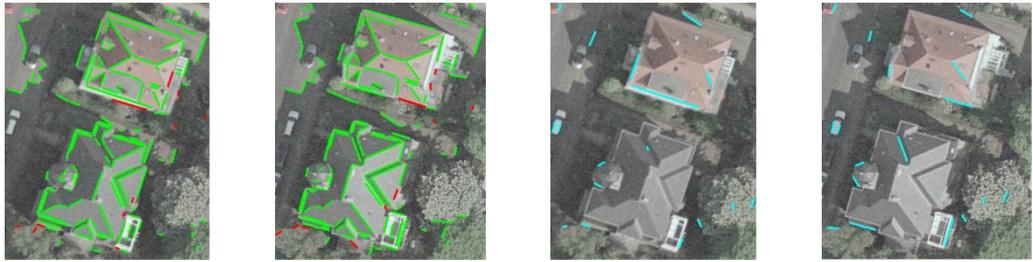
Figure 6.17 (continued)



Test Patch #11 - Matched Lines: 342, Incorrect: 6, Missed: 107



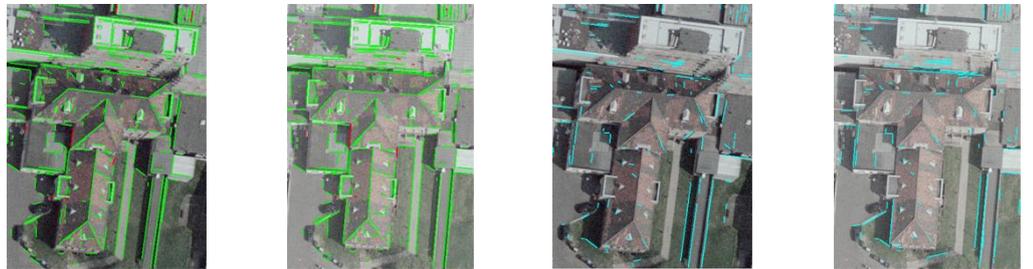
Test Patch #12 - Matched Lines: 92, Incorrect: 3, Missed: 30



Test Patch #13 - Matched Lines: 149, Incorrect: 11, Missed: 17

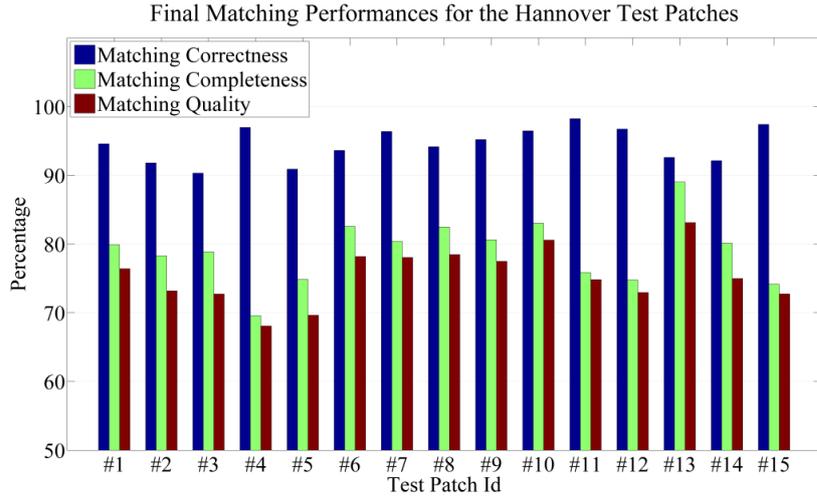


Test Patch #14 - Matched Lines: 140, Incorrect: 11, Missed: 32



Test Patch #15 - Matched Lines: 348, Incorrect: 9, Missed: 118

Figure 6.17 (continued)



Patch Id	Correctness (%)	Completeness (%)	Quality (%)
	$\left[ \frac{TP}{TP + FP} \right]$	$\left[ \frac{TP}{TP + FN} \right]$	$\left[ \frac{TP}{TP + FP + FN} \right]$
#1	94.6	79.9	76.4
#2	91.8	78.3	73.2
#3	90.3	78.9	72.7
#4	96.9	69.6	68.1
#5	90.9	74.9	69.7
#6	93.6	82.6	78.2
#7	96.4	80.4	78.0
#8	94.2	82.5	78.5
#9	95.2	80.6	77.5
#10	96.5	83.0	80.6
#11	98.3	75.8	74.8
#12	96.7	74.8	72.9
#13	92.6	89.0	83.1
#14	92.1	80.1	75.0
#15	97.4	74.2	72.7
<b>Patches 1-6 (<math>\mu \pm \sigma</math>)</b>	<b>93.0 <math>\pm</math> 2.5</b>	<b>77.4 <math>\pm</math> 4.6</b>	<b>73.0 <math>\pm</math> 3.9</b>
<b>Patches 7-15 (<math>\mu \pm \sigma</math>)</b>	<b>95.5 <math>\pm</math> 2.1</b>	<b>80.1 <math>\pm</math> 4.7</b>	<b>77.0 <math>\pm</math> 3.5</b>
<b>Overall (<math>\mu \pm \sigma</math>)</b>	<b>94.5 <math>\pm</math> 2.5</b>	<b>78.9 <math>\pm</math> 4.7</b>	<b>75.4 <math>\pm</math> 4.0</b>

Figure 6.18 The final matching performances of the proposed approach computed for the 15 test patches selected from the Hannover test site.

the test patches revealed results ranging between 69% and 89%. The overall completeness performance of the matching is computed to be 78.9% with a standard deviation of 4.7%. Actually, the average completeness values are approximately 10% less than the average completeness values achieved for the Vaihingen test patches. In fact, this is an expected result, since the images of the Hannover dataset were taken by an analog camera and scanned afterwards. Thus, the quality of the images and the noise level involved has probably affected the final quality of the extracted line segments. As expected, the quality of the line matching also decreased by an amount of 10% in average compared to the Vaihingen results.

If the incorrect line matches of the Hannover test patches are visually examined, similar to the Vaihingen test patches, most of them appear in the occluded parts. On the other hand, in terms of missed matches, we see that the reasons explicitly stated for the Vaihingen patches are also valid for the Hannover test patches. However, in this case, the fragmentation of a single line into multiple line segments due to poor image quality mostly dominated the number of missed matches and seems to be the major reason for the 10% decrease compared to the Vaihingen results. Nevertheless, it is believed that the proposed approach revealed very good line matching results for both the Vaihingen and Hannover test datasets.

### **6.3.3 Comparison to the State-of-the-Art**

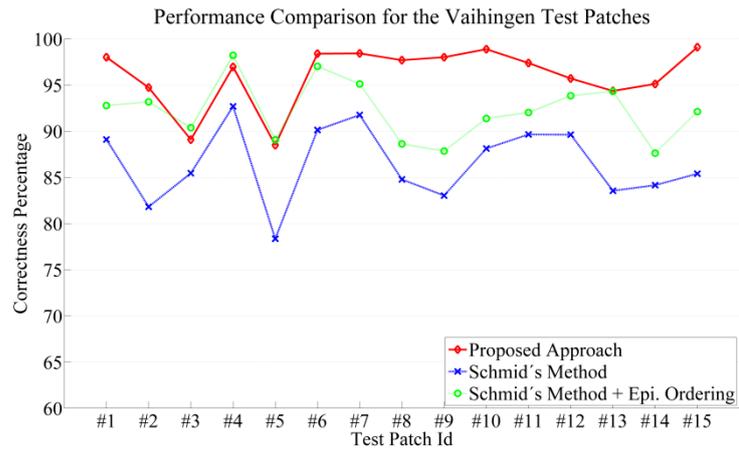
In this section, we compared the results of the proposed approach with the results of the stereo matching approach presented in Schmid and Zisserman (1997). They proposed a line matching algorithm which utilizes direct and warped correlation measures computed around the line neighborhoods. The related MatLab toolbox of the approach can be found in Werner and Zisserman (2002). Here, we also investigated the approach with and without epipolar ordering constraint (Werner, 2002) in order to fully expose the stereo matching performances of their approach.

In order to perform a meaningful comparison, for both approaches, we utilized the same straight line segments extracted in the previous section, and the comparative results in terms of the correctness, completeness and quality levels for the Vaihingen and Hannover test patches are provided (Table 6.4-6.5, Fig. 6.19-6.20). Based on the correctness results given in Fig. 6.19a, for all Vaihingen test patches, the performance of the proposed approach exceeded the performance of the Schmid's method. After imposing the epipolar ordering constraint to the Schmid's method, it is clear that the correctness level reached by the method is significantly increased. However, for most of the test patches, even after applying the constraint, the Schmid's method cannot reach the same correctness level of the proposed approach. If we take into account the completeness values computed for the approaches, first, it is clear that the ordering constraint has no significant effect on the completeness ratios. Besides, the levels of completeness computed for the approaches are also comparable. It is clear that the proposed approach provided better completeness values for ten of the fifteen test patches. If we look in detail the five test patches (#4-6, #14-15) that the proposed approach resulted in lower completeness values with respect to the Schmid's method, it is clear that there is correlation between the total number of line segments that cannot be matched around the side and the corners of the tests patches. Although those line matches are very distinct and clear, as already indicated before, it may not be possible to match those line segments in a pair-wise logic. However, since the Schmid method searches the corresponding line segments independent of line relations, apparently, the method successfully recovered those matches. Nevertheless, note that, the proposed approach provided a total of 3% improvement for the computed completeness values in an overall sense (Table 6.4).

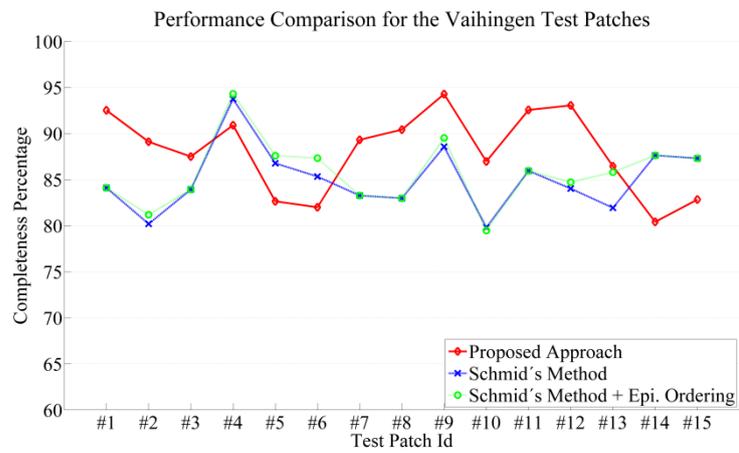
For the Hannover test patches, it is clear that the proposed approach completely outperformed the Schmid's method in terms of the computed correctness values (Fig. 6.20). Once again, it is clear that the epipolar ordering constraint has boosted the stereo matching performance of the Schmid's method. Nonetheless, the

Table 6.4 Comparative matching results of the proposed approach and the Schmid's method for the Vaihingen test patches.

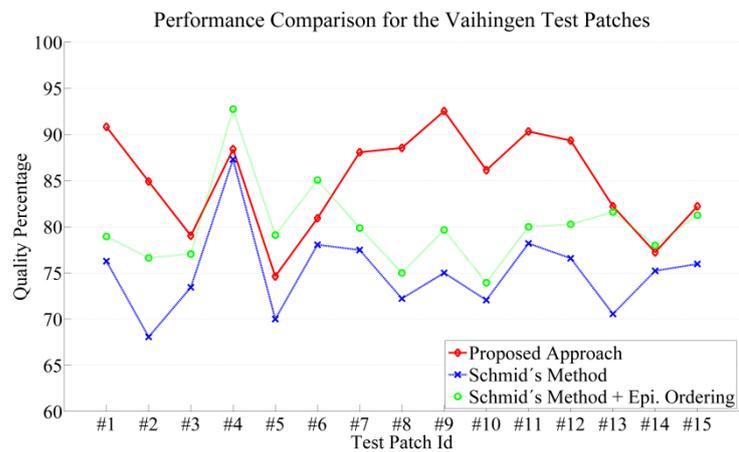
Patch Id	Correctness – Completeness – Quality (%) {TP / FP / FN}		
	Proposed Approach	Schmid's Method	Schmid's Method + Epipolar Ordering
#1	98.0 – 92.5 – 90.8 {99 / 2 / 8}	89.1 – 84.1 – 76.3 {90 / 11 / 17}	92.8 – 84.1 – 78.9 {90 / 7 / 17}
#2	94.7 – 89.1 – 84.9 {90 / 5 / 11}	81.8 – 80.2 – 68.1 {81 / 18 / 20}	93.2 – 81.2 – 76.6 {82 / 6 / 19}
#3	89.1 – 87.5 – 79.0 {49 / 6 / 7}	85.5 – 83.9 – 73.4 {47 / 8 / 9}	90.4 – 83.9 – 77.0 {47 / 5 / 9}
#4	96.9 – 90.9 – 88.4 {160 / 5 / 16}	92.7 – 93.8 – 87.3 {165 / 13 / 11}	98.2 – 94.3 – 92.7 {166 / 3 / 10}
#5	88.5 – 82.6 – 74.6 {100 / 13 / 21}	78.4 – 86.8 – 70.0 {105 / 29 / 16}	89.1 – 87.6 – 79.1 {106 / 13 / 15}
#6	98.4 – 82.0 – 80.9 {123 / 2 / 27}	90.1 – 85.3 – 78.1 {128 / 14 / 22}	97.0 – 87.3 – 85.1 {131 / 4 / 19}
#7	98.4 – 89.3 – 88.1 {251 / 4 / 30}	91.8 – 83.2 – 77.5 {234 / 21 / 47}	95.1 – 83.3 – 79.9 {234 / 12 / 47}
#8	97.7 – 90.4 – 88.5 {85 / 2 / 9}	84.8 – 82.9 – 72.2 {78 / 14 / 16}	88.6 – 82.9 – 75.0 {78 / 10 / 16}
#9	98.0 – 94.2 – 92.5 {99 / 2 / 6}	83.0 – 88.6 – 75.0 {93 / 19 / 12}	87.9 – 89.5 – 79.7 {94 / 13 / 11}
#10	98.9 – 86.9 – 86.1 {267 / 3 / 40}	88.1 – 79.8 – 72.1 {245 / 33 / 62}	91.4 – 79.5 – 73.9 {244 / 23 / 63}
#11	97.4 – 92.6 – 90.3 {112 / 3 / 9}	89.7 – 85.9 – 78.2 {104 / 12 / 17}	92.0 – 85.9 – 80.0 {104 / 9 / 17}
#12	95.7 – 93.1 – 89.3 {134 / 6 / 10}	89.6 – 84.0 – 76.6 {121 / 14 / 23}	93.8 – 84.7 – 80.3 {122 / 8 / 22}
#13	94.4 – 86.5 – 82.2 {134 / 8 / 21}	83.6 – 81.9 – 70.6 {127 / 25 / 28}	94.3 – 85.8 – 81.6 {133 / 8 / 22}
#14	95.1 – 80.4 – 77.2 {78 / 4 / 19}	84.2 – 87.6 – 75.2 {85 / 16 / 12}	87.6 – 87.6 – 77.9 {85 / 12 / 12}
#15	99.1 – 82.8 – 82.2 {111 / 1 / 23}	85.4 – 87.3 – 75.9 {117 / 20 / 17}	92.1 – 87.3 – 81.3 {117 / 10 / 17}
<b>Overall mean</b>	96.0 – 88.1 – 85.0	86.5 – 85.0 – 75.1	92.2 – 85.7 – 79.9



(a)



(b)

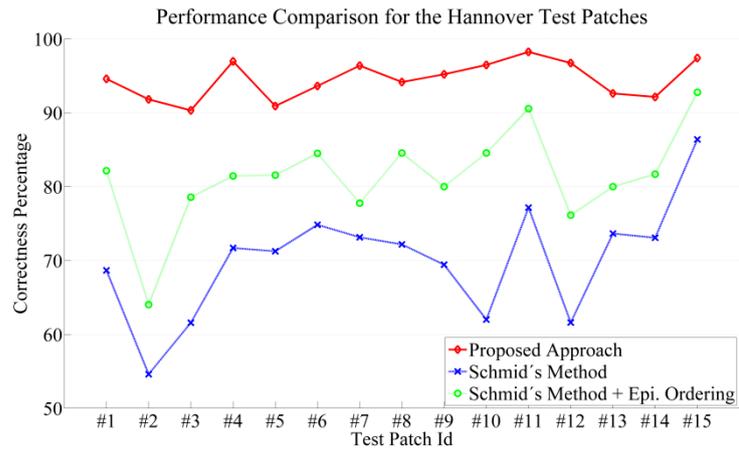


(c)

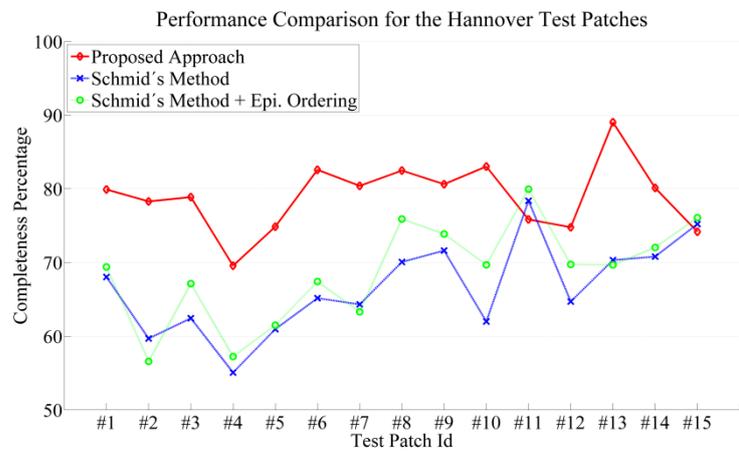
Figure 6.19 Performance comparisons of the proposed approach and the Schmid's method for the Vaihingen test patches.

Table 6.5 Comparative matching results of the proposed approach and the Schmid's method for the Hannover test patches.

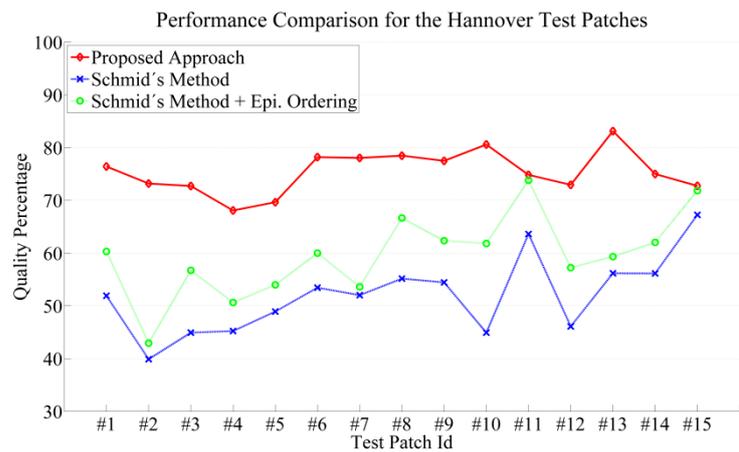
Patch Id	Correctness – Completeness – Quality (%) {TP / FP / FN}		
	Proposed Approach	Schmid's Method	Schmid's Method + Epipolar Ordering
#1	94.6 – 79.9 – 76.4 {185 / 10 / 44}	68.7 – 68.0 – 51.9 {217 / 68 / 70}	82.2 – 69.4 – 60.3 {185 / 33 / 67}
#2	91.8 – 78.3 – 73.2 {110 / 9 / 28}	54.6 – 59.7 – 39.9 {141 / 64 / 52}	64.0 – 56.6 – 42.9 {114 / 41 / 56}
#3	90.3 – 78.9 – 72.7 {186 / 18 / 45}	61.6 – 62.4 – 44.9 {216 / 83 / 80}	78.6 – 67.1 – 56.7 {182 / 39 / 70}
#4	96.9 – 69.6 – 68.1 {99 / 3 / 42}	71.7 – 55.1 – 45.2 {106 / 30 / 62}	81.4 – 57.3 – 50.6 {97 / 18 / 59}
#5	90.9 – 74.9 – 69.7 {154 / 14 / 47}	71.3 – 60.9 – 48.9 {160 / 46 / 73}	81.6 – 61.5 – 54.0 {141 / 26 / 72}
#6	93.6 – 82.6 – 78.2 {157 / 10 / 31}	74.8 – 65.2 – 53.5 {155 / 39 / 62}	84.5 – 67.4 – 60.0 {142 / 22 / 58}
#7	96.4 – 80.4 – 78.0 {166 / 6 / 39}	73.1 – 64.3 – 52.0 {175 / 47 / 71}	77.8 – 63.3 – 53.6 {162 / 36 / 73}
#8	94.2 – 82.5 – 78.5 {120 / 7 / 24}	72.2 – 70.1 – 55.2 {133 / 37 / 41}	84.6 – 75.9 – 66.7 {123 / 19 / 33}
#9	95.2 – 80.6 – 77.5 {188 / 9 / 43}	69.4 – 71.6 – 54.5 {229 / 70 / 63}	80.0 – 73.9 – 62.4 {205 / 41 / 58}
#10	96.5 – 83.0 – 80.6 {142 / 5 / 28}	62.0 – 62.0 – 44.9 {150 / 42 / 57}	84.6 – 69.7 – 61.8 {136 / 21 / 50}
#11	98.2 – 75.8 – 74.8 {342 / 6 / 107}	77.2 – 78.4 – 63.6 {451 / 103 / 96}	90.6 – 79.9 – 73.8 {392 / 37 / 89}
#12	96.7 – 74.8 – 72.9 {92 / 3 / 30}	61.6 – 64.7 – 46.1 {125 / 48 / 42}	76.2 – 69.8 – 57.2 {109 / 26 / 36}
#13	92.6 – 89.0 – 83.1 {149 / 11 / 17}	73.6 – 70.3 – 56.2 {148 / 39 / 46}	80.0 – 69.7 – 59.3 {135 / 27 / 47}
#14	92.1 – 80.1 – 75.0 {140 / 11 / 32}	73.1 – 70.8 – 56.2 {156 / 42 / 47}	81.7 – 72.1 – 62.0 {142 / 26 / 45}
#15	97.4 – 74.2 – 72.7 {348 / 9 / 118}	86.4 – 75.2 – 67.3 {397 / 54 / 113}	92.8 – 76.1 – 71.8 {374 / 27 / 109}
<b>Overall mean</b>	94.5 – 78.9 – 75.4	70.1 – 66.6 – 52.0	81.4 – 68.6 – 59.6



(a)



(b)



(c)

Figure 6.20 Performance comparisons of the proposed approach and the Schmid's method for the Hannover test patches.

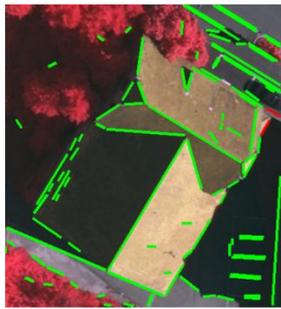
proposed approach clearly provides better correctness levels with an overall performance improvement of 13% (Table 6.5). Similarly, the proposed approach provided an overall completeness performance improvement of around 10% with a final overall quality improvement of about 15% (Table 6.5). Thus, if the results are taken as a whole, we can easily conclude that the proposed approach presented highly promising line matching results for the selected test patches from the Vaihingen and Hannover test datasets.

It is clear that one of the main disadvantages of the proposed approach is the algorithm complexity due to the pair-wise nature. Therefore, as a final comment, compared to the Schmid's method, the proposed approach requires a significant amount of time, especially for large areas where an enormous number of line segments are observed.

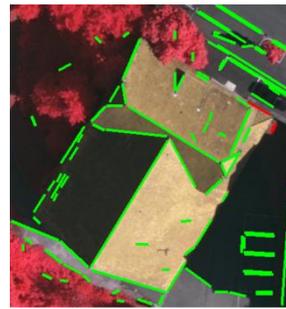
#### **6.4 The Results and Discussion of the Proposed Line Reconstruction Approach**

The performance assessment of the reconstruction is performed for three image patches selected from the test site Vaihingen (Fig. 6.21). Those test patches are intentionally selected based on their special characteristics of which a number of line segments in each patch are observed in a (nearly) aligned condition with the epipolar line (from left to right). The matching results revealed that, for all three test patches, very high correctness levels are achieved. The correct and false matches are explicitly illustrated in Fig. 6.21. The matching results indicate that the proposed line matching approach remains robust for the image patches that contain a significant number of line segments that are (nearly) parallel to the epipolar line. For example, for the third image patch, approximately half of the extracted line segments are nearly-aligned with the epipolar line (Fig. 6.21).

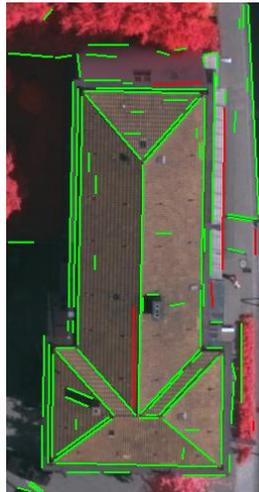
Despite the successful line matching performances, the classical direct construction method produced dramatic reconstruction problems for the lines that



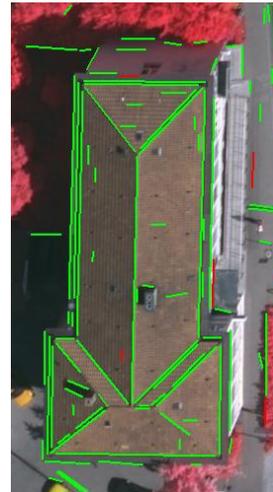
(a)



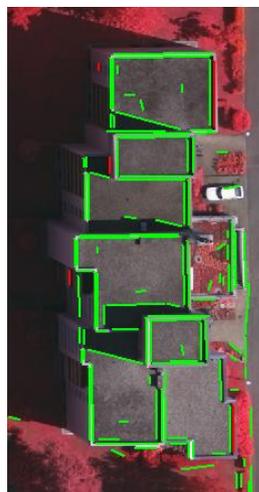
(b)



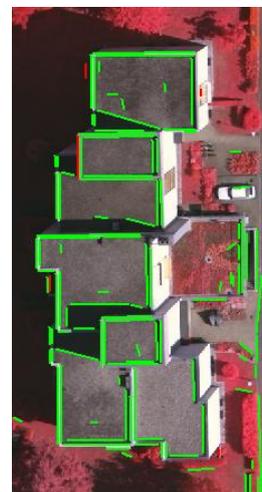
(c)



(d)



(e)



(f)

Figure 6.21 Test patches for the evaluation of the proposed reconstruction approach. Left (a-c-e) and right (b-d-f) stereo images. Correct and false matches are shown in green and red colors, respectively.

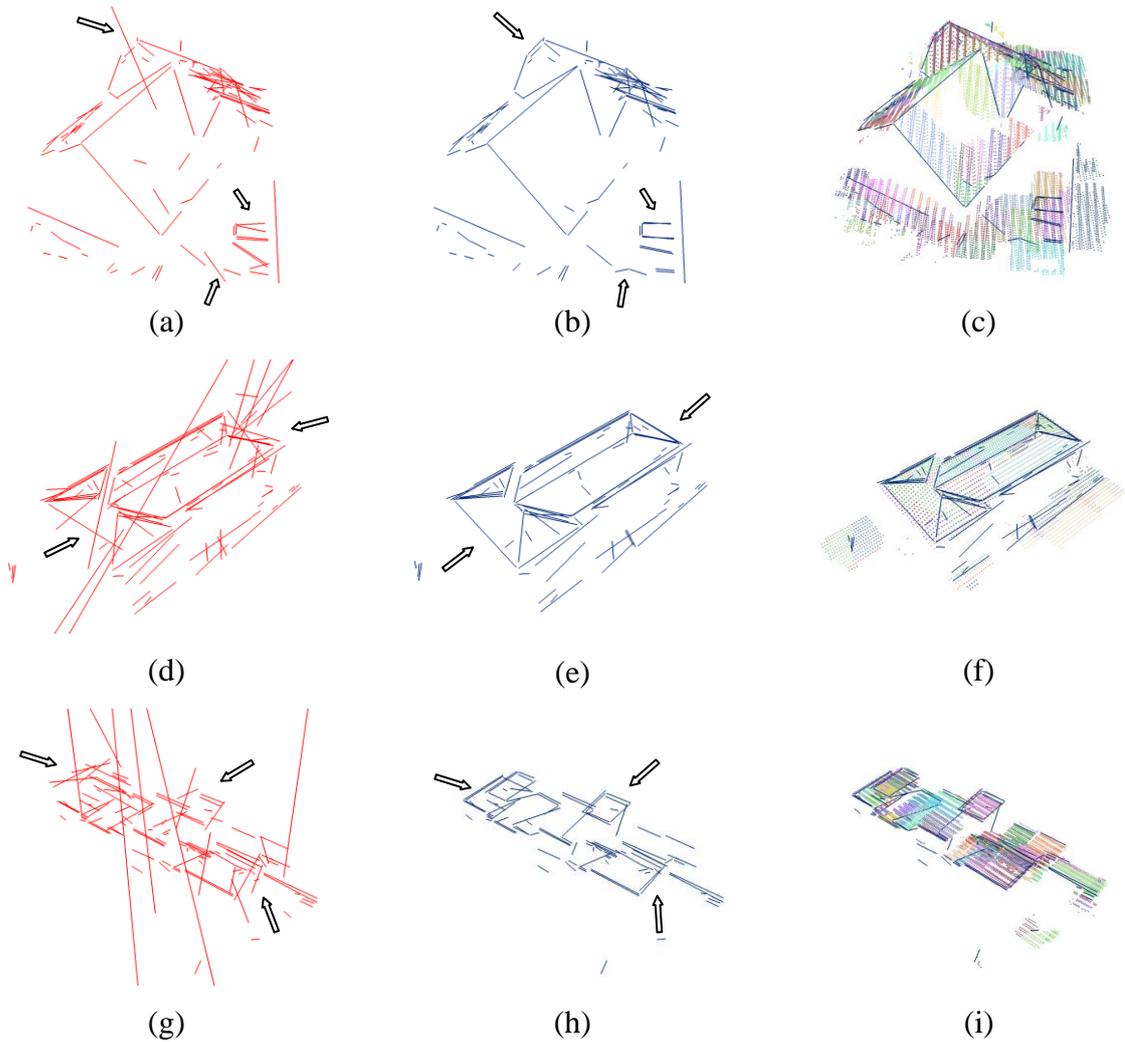


Figure 6.22 3D line segments generated with the method of direct construction (a-d-g) and the proposed reconstruction approach (b-e-h). Block arrows point parts where some of the critical improvements are observed. The LIDAR point cloud overlaid with the lines reconstructed with the proposed approach (c-f-i). Each color in the point cloud belongs to the automatically extracted 3D planes in the vicinity of each line.

are nearly-aligned with the epipolar line (Fig. 6.22 a-d-g). It is clear from those figures that the reconstruction results of the line segments that are nearly-aligned with the epipolar line are extremely defective and irrelevant. On the other hand, the proposed reconstruction approach successfully recovered most of those

Table 6.6 The computed RMS distances for the method of Direct Construction

Patch ID	RMS Average Distance (m)		
	Not-Aligned	Nearly-Aligned	Total
#1	0.152	1.041	0.495
#2	0.173	4.451	2.240
#3	0.194	6.278	3.768

Table 6.7 The computed RMS distances for the proposed approach

Patch ID	RMS Average Distance (m)		
	Not-Aligned	Nearly-Aligned	Total
#1	0.152	0.357	0.204
#2	0.173	0.196	0.179
#3	0.194	0.459	0.275

problematic cases (Fig. 6.22 b-e-h). Although we believe that the level of improvement is visually apparent, we also evaluated the accuracy of the reconstructed line segments by comparing them to LIDAR data. In order to compare the reconstructed lines, we automatically extracted 3D planes from the point cloud in the vicinity of each line (Fig. 6.22 c-f-i). Thereafter, we determined the line's average orthogonal distance from its neighbouring planes and used these distances to compute the RMS average distance between the reconstructed lines and the LIDAR planes. The quantitative results of the direct construction and the proposed reconstruction approach for each patch are provided in Tables 6.6 and 6.7, respectively. As expected, the method of direct construction produced reasonable RMS distances ( $\approx 2$  pixels) for the line segments that are not aligned with the epipolar line. On the other hand, large RMS distances ( $> 1$  m) are

inevitable for the nearly-aligned cases, which also reduce the overall performance considerably. In contrast to the results of the direct construction, for each test patch, our approach led to massive RMS improvements after the reconstruction of the nearly-aligned line segments. Not surprisingly, in each case, this achievement reflects to the overall RMS performances as well. More significantly, for the second patch, the final RMS distances of the nearly-aligned cases (0.196 m) are almost at the same level of the not-aligned cases (0.173 m). This fact also confirms the quality level that can be reached by the proposed reconstruction approach.

Based on those results, we can easily conclude that, the proposed approach can produce highly promising reconstruction results for the line segments that are (nearly) aligned with the epipolar line. Indeed, the final improvements are massive in terms of RMS performance; if we think that the results of the previous approaches that rely on just direct construction consistently result in large RMS errors ( $> 1$  m). It is also obvious that the selected 3D point entities ( $\mathbf{X}_i$ ) during the estimation process determine the final height quality of the 3D line estimated. In a worst-case scenario in which only a single 3D point entity is available, our method will not perform worse than the standard direct construction.

In Fig. 6.23 and 6.24, the final reconstruction results for the entire Vaihingen and Hannover patches are visualized.

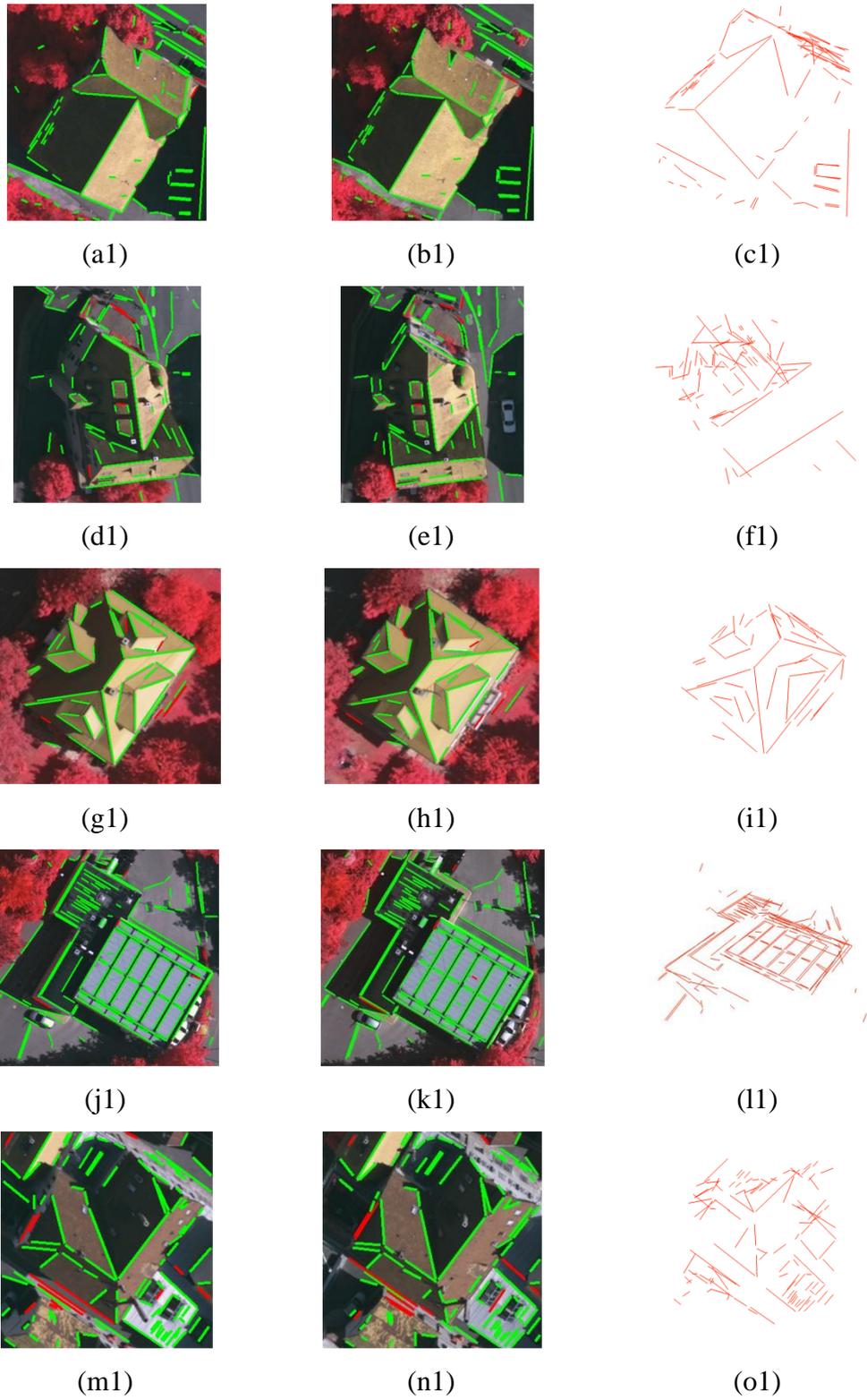
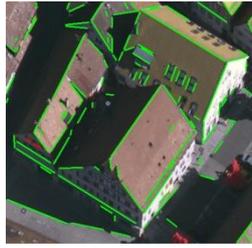
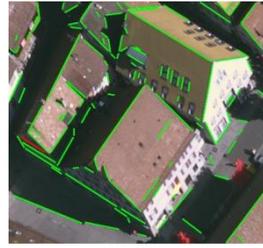


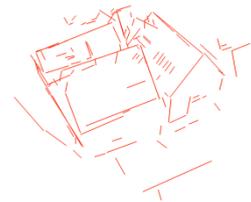
Figure 6.23 The final reconstruction results for the Vaihingen test patches



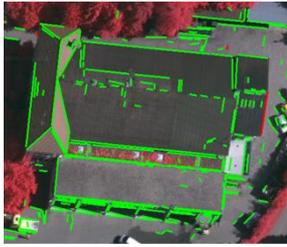
(a2)



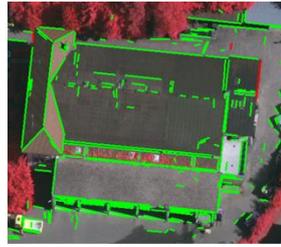
(b2)



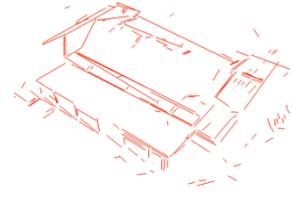
(c2)



(d2)



(e2)



(f2)



(g2)



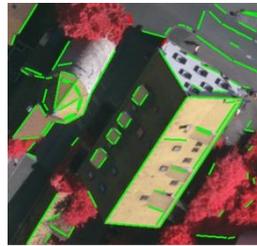
(h2)



(i2)



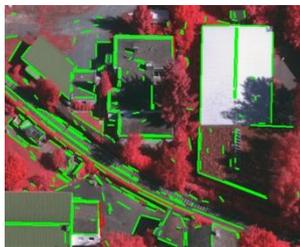
(j2)



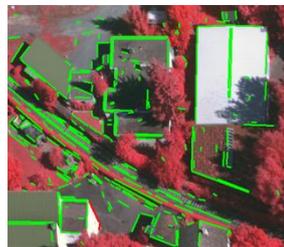
(k2)



(l2)



(m2)

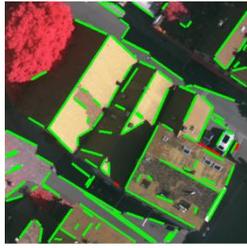


(n2)

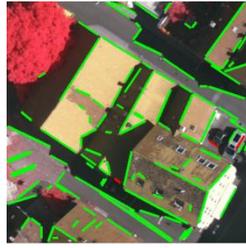


(o2)

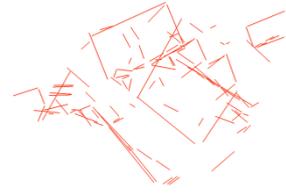
Figure 6.23 (continued)



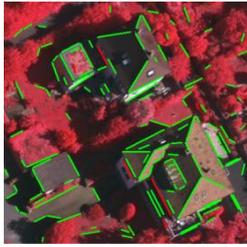
(a3)



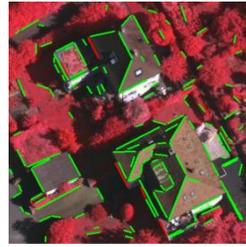
(b3)



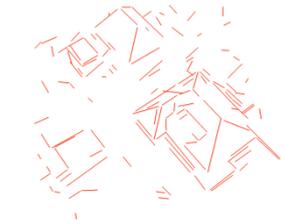
(c3)



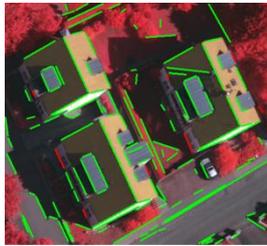
(d3)



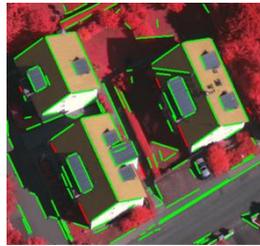
(e3)



(f3)



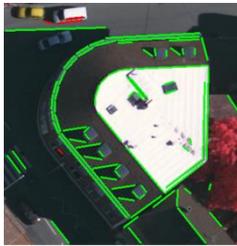
(g3)



(h3)



(i3)



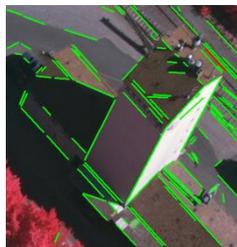
(j3)



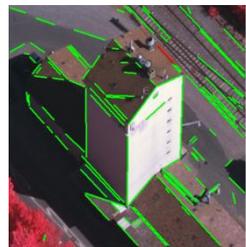
(k3)



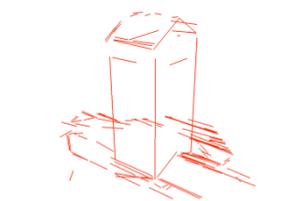
(l3)



(m3)



(n3)



(o3)

Figure 6.23 (continued)

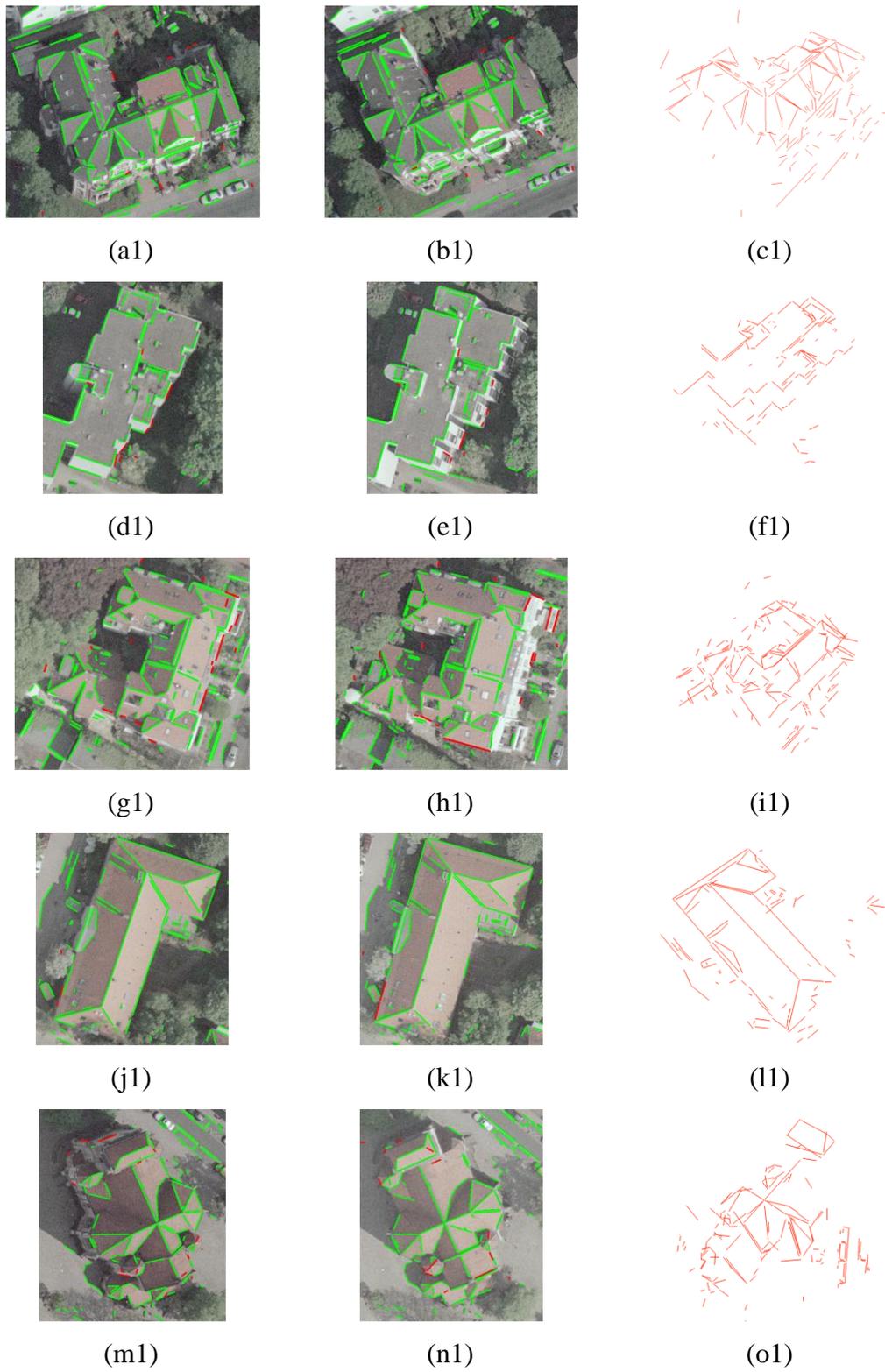
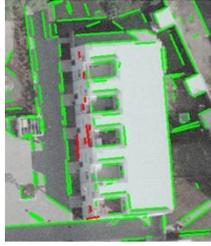
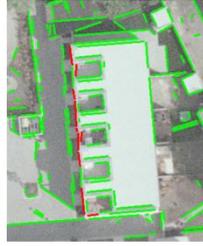


Figure 6.24 The final reconstruction results for the Hannover test patches



(a2)



(b2)



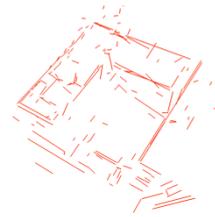
(c2)



(d2)



(e2)



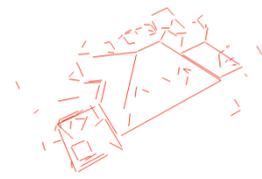
(f2)



(g2)



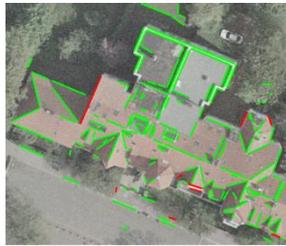
(h2)



(i2)



(j2)



(k2)



(l2)



(m2)

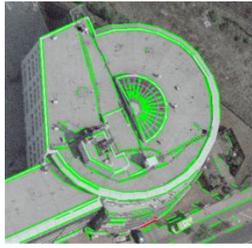


(n2)

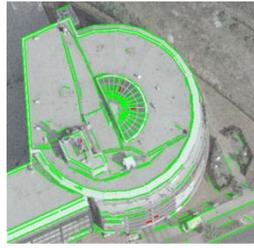


(o2)

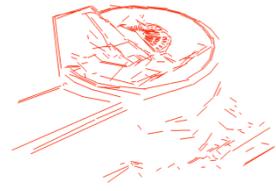
Figure 6.24 (continued)



(a3)



(b3)



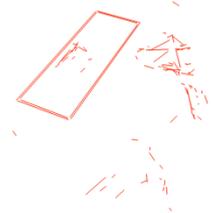
(c3)



(d3)



(e3)



(f3)



(g3)



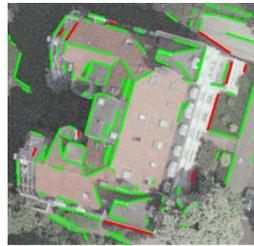
(h3)



(i3)



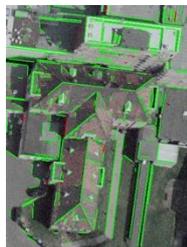
(j3)



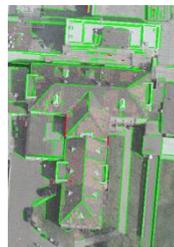
(k3)



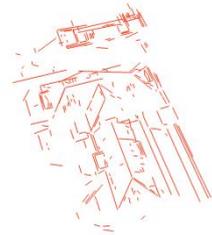
(l3)



(m3)



(n3)



(o3)

Figure 6.24 (continued)

## CHAPTER 7

### CONCLUSIONS AND RECOMMENDATIONS

In this chapter, the conclusions derived from the developed approaches are stated and the recommendations regarding to possible further studies are given.

#### 7.1 Conclusions

The following conclusions are reached from the results achieved for the proposed line extraction, stereo line matching, and stereo reconstruction approaches:

- The proposed line extraction approach takes full advantage of the existing multispectral information in aerial images all over the steps especially for the pre-processing and edge detection. Thus, even object boundaries that show only a very slight color difference could be detected.
- With the improvements performed to the straight edge detector, the straight line extraction algorithm works quite robust, even for the areas where an enormous number of edges were found. This offers an opportunity to concentrate on not only to the major components of the building objects but also to certain details such as dormers, solar panels, etc.
- The line matching results computed indicate that, the proposed line matching approach provides quite promising results. For all test patches

selected from the Vaihingen and Hannover test sites, the overall correctness and completeness values are computed to be 95.2% and 83.5%, respectively.

- For the Vaihingen test patches selected, the proposed line matching approach revealed very high rates of correctness values ranging between 88% and 99%. Among all 15 test patches, 13 of them have reached correctness levels of over 94%, thus the overall correctness performance of matching is computed to be 96% with a standard deviation of 3.3%. The completeness values of the line matches for those test patches also revealed remarkable results ranging between 80% and 94%. The overall completeness value is computed to be 88.1% with a standard deviation of 4.4%.
- According to the matching results of the Hannover test patches, the proposed line matching approach revealed quite satisfactory correctness values ranging between 90% and 98%. The overall correctness performance of the matching is computed to be 94.5% with a standard deviation of 2.5%. On the other hand, the completeness levels of the line matches for the test patches are computed to be in the range of 69% and 89%. The overall completeness performance of the matching is computed to be 78.9% with a standard deviation of 4.7%.
- According to the PDFs computed for the developed pair-wise constraints, the separation of the entire correct pair-wise matches from the incorrect ones is not possible with just specifying pair-wise constraints. However, those pair-wise constraints enable us to develop a local matching support during the precise matching, and to decrease the level of complexity involved during the pair-wise line matching. According to the results

achieved, the total number of line matching possibilities can be reduced around 80% after imposing the pair-wise constraints.

- It is found from the PDFs computed for the Vaihingen and Hannover test patches that the proposed line-based Daisy ( $Sim_D$ ) measure are very useful to identify the correct and incorrect matches during the precise matching stage. Furthermore, the proposed Daisy dissimilarity NNDR ( $1 - Sim_D$ ) metric can have possibility to remove a large number of incorrect line matches beforehand without removing any of the correct matches.
- The exclusive line local matching support that is proposed by the normalized Redundancy ( $Sim_R$ ) measure is also found to be highly discriminative between the correct and incorrect line matches. As strongly anticipated, the incorrect line matches provided very low similarity values for the Redundancy measure; whereas it is proved that most of the correct line matches have significantly larger local matching support. Thus, this information gives an excellent opportunity to separate the correct line matches from the incorrect ones during matching of line segments.
- For the final step of the precise matching, it is proved that the proposed hysteresis like thresholding can eliminate most of the remaining incorrect line matches while keeping the desired levels of matching correctness and completeness. This turns out to be one of the very strong aspects of the proposed approach, since all previous approaches relies on a final thresholding based on a single measure; thus, for those approaches, it may not be possible to provide a good balance between the levels of correctness and completeness.
- In this study, it is also verified that the entire threshold values utilized during the line matching stage can be fixed for the aerial stereo image pairs that are acquired in a consecutive order. In addition, the threshold

values do not depend on the type of the aerial image used (analogue or digital) and can also be applied successfully for the aerial images with standard forward overlaps of 60% or more.

- The line matching results computed for the Hannover test sites revealed that the quality of the line matching is decreased for an amount of 10% in average compared to the Vaihingen results. In fact, this is an expected result, since the images of the Hannover dataset were taken by an analogue camera and scanned afterwards. Thus, the quality of the images and the noise level involved has probably affected the line matching performance. As a conclusion, the approach can provide better line matching performances for the aerial images acquired by digital cameras.
- Comparative results between the proposed line matching approach with the state-of-the-art line matching approach presented in Schmid and Zisserman (1997) revealed that the proposed approach clearly outperformed the Schmid's method in all aspects. It is also found that the proposed approach can also provide better performances in an overall sense compared to the Schmid's method which is enhanced with the constraint of epipolar ordering.
- According to the analyses of the incorrect line matches, it is clear that the occlusion is still the major reason of the mismatches for the proposed line matching approach. Nevertheless, the computed correctness levels for the test patches prove that, since the proposed line matching approach handles the matching by means of line pairs; apparently, this configuration is more robust and efficient to handle the occlusion phenomena.
- The reconstruction results provided by the method of direct construction revealed gross RMS errors ( $> 1$  m) for the line segments that are nearly-aligned with the epipolar line. It can be easily concluded that the

reconstructed line segments by means of direct construction are not reliable and useful for those ill-posed cases.

- The proposed stereo reconstruction approach produced highly promising results for the line segments that are nearly-aligned with the epipolar line. The numerical and visual evaluations revealed that the proposed approach completely outperformed the method of direct construction, and provided substantial improvements for the final height values of the reconstructed line segments. Thus, as a conclusion, we have a possibility to reconstruct those ill-posed cases with promising final accuracies.

## **7.2 Recommendations**

The followings are recommended for further studies:

- The matching of more than two images can provide solutions for the occlusion phenomena. Additionally, higher reconstruction performances can be achieved through the intersection of more than two projection planes. Inspired by Baillard et al. (1999), an obvious future work is to improve and extend the current stereo approach to the cases in which multiple aerial images are available.
- It is clear that one of the main disadvantages of the proposed approach is the algorithm complexity due to the pair-wise nature. Therefore, how to provide a compromise between the algorithm complexity and matching efficiency constitutes another topic for further development.
- In principle, the proposed line matching and reconstruction approaches can be further extended to other image sources such as linear array sensors and close-range images. Extending the approach based on the geometry of linear array sensors will give a chance to reconstruct the line segments for

UHR aerial linear array images and as well as forthcoming very high resolution satellite images that are planned to have image resolutions of around 30 cm (such as WorldView-3).

- Integrating coarse DSMs/DTMs generated from dense stereo image matching algorithms into the line matching stage may have possibility to further reduce the number of mismatches and to increase the speed of the processing. A future work may include experimental tests based on specific DSMs/DTMs generated from stereo aerial images.
- The reconstructed line segments by the proposed approach can be used in different applications such as the complete reconstruction of the building and road features, improving the quality of the existing dense DSMs etc. After the complete reconstruction, a different future application area of this work also involve the generation of virtual environments with promising level-of-details.

## REFERENCES

Abramson, S.B., Schowengerdt, R.A., 1993. Evaluation of Edge-Preserving Smoothing Filters for Digital Image Mapping. *International Journal of Photogrammetry and Remote Sensing*, Vol. 48, pp. 2-17.

Atalay, V., and Yilmaz, M.U., 1998. A Matching Algorithm Based On Linear Features. *Pattern Recognition Letters*, Vol. 19, pp. 857-867.

Baillard, C., and Dissard, O., 2000. A Stereo Matching Algorithm for Urban Digital Elevation Models. *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 9, pp. 1119-1128.

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic Line Matching and 3D Reconstruction of Buildings from Multiple Views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5, pp. 69-80.

Ballard, D., 1981. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, Vol. 13, No. 2, pp. 111-122.

Bay, H., Tuytelaars, T., and Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: European Conference on Computer Vision.

Beder, C., 2004. A Unified Framework for the Automatic Matching of Points and Lines in Multiple Oriented Images. In: Proceedings of XXth ISPRS Congress. Istanbul, Turkey, pp. 1109-1113.

Ben-Tzvi, D., Sandler, M.B., 1990. A Combinatorial Hough Transform. *Pattern Recognition Letters*, Vol. 11, No. 3, pp. 167-174.

Bignone, F., Henricsson, O., Fua, P. and Stricker, M., 1996. Automatic Extraction of Generic House Roofs From High Resolution Aerial Imagery. In: European Conference on Computer Vision, pp. 85-96.

Bilen, B., 2004. Model Based Building Extraction from High Resolution Aerial Images. M.Sci. Thesis, Middle East Technical University.

Burns, J.B., Hanson, A.R., 1986. Extraction Straight Lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 4, pp. 425-456.

Canny, J., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, pp. 679-698.

Chehata, N., Deseilligny, M.P., Jung, F., and Stamon, G., 2002. Extraction of 3D Primitives from Stereopairs of Satellite Images for Automatic Reconstruction of Buildings. In: IAPR Workshop on Machine Vision Applications, Nara- ken New Public Hall, Nara, Japan.

Cheng, Y.Q., Wang, X.G., Collins, R.T., Riseman, E.M., and Hanson, A.R., 2001. Three-Dimensional Reconstruction of Points and Lines with Unknown Correspondence across Images. *International Journal of Computer Vision*, Vol. 45, No. 2, pp. 29-156.

CIE, 1976. International Commission on Illumination.

Cochran, S.D., and Medioni, G., 1992. 3D Surface Description from Binocular Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14., No. 10, pp. 981-994.

Collins, R.T., Jaynes, C.O., Cheng, Y.Q., Wang, X., Stolle, F., Riseman, E.M., and Hanson, A.R., 1998. The Ascender System: Automated Site Modeling from Multiple Aerial Images. *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 143-162.

Cord, M., and Declercq, D., 2001. Three-Dimensional Building Detection and Modeling using a Statistical Approach. *IEEE Transactions on Image Processing*, Vol. 10, No. 5, pp. 715-723.

Cramer, M., 2005. Digital Airborne Cameras - Status and Future. In: Proceedings of ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information, 17-20 May, Hannover, Germany.

Cramer, M., 2010. The DGPF Test on Digital Aerial Camera Evaluation – Overview and Test Design. *Photogrammetrie–Fernerkundung–Geoinformation*, Vol. 2, pp. 73-82.

Deriche, R., 1987. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision*, Vol. 1, pp. 167-187.

Deriche, R., Vaillant, R., and Faugeras, O., 1991. From Noisy Edge Points to 3D Reconstruction of a Scene: A Robust Approach and Its Uncertainty Analysis. In: Seventh Scandinavian Conference on Image Analysis, Aalborg, Denmark, pp. 225-232.

Dhond, U., and Aggarwal, J.K., 1989. Structure from Stereo: A Review. *IEEE Transactions on System, Man, and Cybernetics*, SMC-19, No. 6, pp. 1489-1510 .

Di Zeno, S., 1986. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, Vol. 33, No. 1, pp. 116-125.

Elaksher, A.F., Bethel, J.S., and Mikhail, E.M., 2003. Roof Boundary Extraction using Multiple Images. *Photogrammetric Record*, Vol. 18, No. 101, pp. 27-40.

Faugeras, O., 1993. Three-Dimensional Computer Vision, a Geometric Viewpoint. MIT Press.

Förstner, W., 1994. A framework for low level feature extraction. In: ECCV '94: Proceedings of the third European conference on Computer Vision (Vol. II) (1994), pp. 383-394.

Förstner, W., 2002. Computer Vision and Photogrammetry - Mutual Questions: Geometry, Statistics and Cognition. In: Bildtechnik/Image Science, Swedish Society for Photogrammetry and Remote Sensing. pp. 151-164.

Förstner, W., 2004. Projective Geometry for Photogrammetric Orientation Procedures II. In: Tutorial Notes From the Tutorial Held at the ISPRS Congress Istanbul. Istanbul, Turkey.

Förstner, W., 2005. Uncertainty and Projective Geometry. In: Bayro Corrochano, Eduardo (Hg.): Handbook of Geometric Computing. 2005, pp. 493-535.

Förstner, W., 2009. Computer Vision and Remote Sensing: Lessons Learned. Dieter Fritsch (Ed.), Photogrammetric Week 2009, pp. 241-249.

Förstner, W., 2010. Minimal Representations for Uncertainty and Estimation in Projective Spaces. In: Proceedings of the Asian Conference on Computer Vision. Queenstown, New Zealand, pp. 619-633, Part II.

Förstner, W., Brunn, A., and Heuel, S., 2000. Statistically Testing Uncertain Geometric Relations. In: G. Sommer, N. Krüger, and Ch. Perwass, editors, Mustererkennung, pp. 17-26. DAGM, Springer.

Fraser, C.S., 1997. Digital Camera Self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 149-159.

Fritsch, D., Stallmann, D., 2000. Rigorous Photogrammetric Processing of High Resolution Satellite Imagery. ISPRS Congress, vol. XXXIII.

Gevers, T., 2006. Color Feature Detection: An Overview Color Image Processing: Methods and Applications. CRC Press.

Greenfeld, J.S., and A.F. Schenk, 1989. Experiments with Edge-Based Stereo Matching. *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 12, 1771-1777.

Güler, M. A., 2004. Detection of Earthquake Damaged Buildings from Post-Event Photographs using Perceptual Grouping. M.Sci. Thesis, Middle East Technical University.

Guru, D.S., Shekar, B.H., and Nagabhushan, P., 2004. A Simple and Robust Line Detection Algorithm Based On Small Eigenvalue Analysis. *Pattern Recognition Letters*, Vol. 25, pp. 1-13.

Haala, N., Hastedt, H., Wolff, K., Ressler, C., and Baltrusch, S., 2010. Digital Photogrammetric Camera Evaluation - Generation of Digital Elevation Models. *Photogrammetrie–Fernerkundung–Geoinformation*, Vol. 2, pp. 99-115.

Habib, A.F., Morgan, M., Lee, Y.R., 2002. Bundle Adjustment with Self-Calibration using Straight Lines. *Photogrammetric Record*, Vol. 17, pp. 635-650.

Habib, A.F., Zhai, R., and Kim, C., 2010. Generation of Complex Polyhedral Building Models by Integrating Stereo Aerial Imagery and LIDAR Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 76, No. 5, pp. 609-623.

Hartley, R., and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Heipke, C., 1996. Overview of Image Matching Techniques. In: OEEPE Workshop on the Application of Digital Photogrammetric Workstations, Lausanne, Switzerland, March, 1996.

Heipke, C., 1997. Automation of Interior, Relative, and Absolute Orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 1-19.

Henricsson, O., 1998. The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction. *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 163-184.

Henricsson, O., and Baltsavias, E., 1997. 3-D Building Reconstruction with ARUBA: A Qualitative and Quantitative Evaluation. Automatic Extraction of Man-Made Objects from Aerial and Space Images II (A. Grün, E. Baltsavias, and O. Henricsson, Eds.), Birkhauser Verlag, Basel, pp. 65–76.

Herman, M., and Kanade, T., 1986. Incremental Reconstruction of 3D Scenes from Multiple, Complex Images. *Artificial Intelligence*, Vol. 30, pp. 289-341.

Heuel, S., 2001. Points, Lines and Planes and their Optimal Estimation. In: Proceedings of the DAGM, pp. 92-99.

Heuel, S., 2004. Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction. First edition, Springer.

Heuel, S., and Förstner, W., 2001. Matching, Reconstructing and Grouping 3D Lines from Multiple Views Using Uncertain Projective Geometry. In: Computer Vision and Pattern Recognition, IEEE.

Hoff, W., and Ahuja, N., 1989. Surfaces from Stereo: Integrating Feature Matching Disparity Estimation, and Contour Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 2, pp. 121-136.

Hough, P.V.C., 1962. Method and Means for Recognizing Complex Patterns. US Patent No. 3069654.

Huertas, A., and Nevatia, R., 1988. Detecting Buildings in Aerial Images. *Computer Vision, Graphics, and Image Processing*, Vol. 41, pp. 131-152.

Illingworth, J., Kittler, J., 1987. The adaptive Hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, pp. 690-698.

Jacobsen, K., 2007. Geometry and Information Contents of Large Size Digital Frame Cameras. In: IAPRS XXXVI. Band 1/W51. Hannover.

Jacobsen, K., 2008. Tells the number of pixels the truth? - Effective Resolution of Large Size Digital Frame Cameras. In: American Society of Photogrammetry and Remote Sensing, Annual Convention, Portland.

Jacobsen, K., 2010. Development of Digital Aerial Cameras. In: ISPRS Istanbul Workshop 2010 on Modeling of Optical Airborne and Spaceborne Sensors, WG I/4, Oct. 11-13, In: IAPRS Vol. XXXVIII, part 1/W17.

Jaynes, C., Riseman, E., and Hanson, A., 2003. Recognition and Reconstruction of Buildings from Multiple Aerial Images. *Computer Vision and Image Understanding*, Vol. 82, No. 1, pp. 68-98.

Jordan, M., Cocquerez, J.P., 1995. Three-dimensional description of scenes observed in aerial photography. *Pattern Recognition*, Vol. 28, No: 7, pp. 931-947.

Jung, F., and Paparoditis, N., 2003. Extracting 3D Free-Form Surface Boundaries of Man-Made Objects from Multiple Calibrated Images: A Robust, Accurate and High Resolving Power Edgel Matching and Chaining Approach. In: International Archives of Photogrammetry and Remote Sensing, 34 (Part 3/W8).

Jung, F., Tollu, V., and Paparoditis, N., 2002. Extracting 3d Edgel Hypotheses from Multiple Calibrated Images: A Step towards the Reconstruction of Curved and Straight Object Boundary Lines. In: Proceedings of Photogrammetric Computer Vision (Vol B), Graz, pp. 100-104.

Kim, Z., and Nevatia, R., 1999. Uncertain Reasoning and Learning for Feature Grouping. *Computer Vision and Image Understanding*, Vol. 76, No. 3, pp. 278-288.

Kim, Z., and Nevatia, R., 2004. Automatic Description of Complex Buildings from Multiple Images. *Computer Vision and Image Understanding*, Vol. 96, No. 1, pp. 60-95.

Koc San, D., 2009. Approaches for Automatic Urban Building Extraction and Updating From High Resolution Satellite Imagery. PhD Thesis, Middle East Technical University.

Koch, K.R., 1999. Parameter Estimation and Hypothesis Testing in Linear Models. Springer.

Koschan, A., and Abidi, M., 2005. Detection and Classification of Edges in Color Images. *IEEE Signal Processing Magazine*, Special Issue on Color Image Processing, Vol. 22, No. 1, pp. 64-73.

Kraus, K., 1993. Photogrammetry, Volume 1.

Ladstädter, R., Gruber, M., Wiechert, A., 2010. Monolithic Stitching: One sensor geometry for multiple sensor camera. In: American Society of Photogrammetry and Remote Sensing, Annual Conference.

Leberl, F., Perko, R., Gruber, M., Ponticelli, M., 2003. The UltraCam Large Format Digital Aerial Camera System. In: Proceedings of American Society of Photogrammetry and Remote Sensing Annual Convention, Anchorage, Alaska.

Lee, Y., Koo, H., and Jeong, C., 2006. A Straight Line Detection Using Principal Component Analysis. *Pattern Recognition Letters*, Vol. 27, pp. 1744-1754.

Leloglou, U.M., 2001. Direct Surface Reconstruction from Multiple Views by Surface Growing. PhD Thesis, Middle East Technical University, Ankara.

Li, H., Lavin, M.A., Le Master, R.J., 1986. Fast Hough Transform: A Hierarchical Approach. *Computer Vision Graphics Image Processing*. Vol. 36, pp. 139–161.

Lillesand, T.M., Kiefer, R.W., and Chipman, J.W., 2005. Remote Sensing and Image Interpretation. Fifth Edition, John Wiley and Sons.

Lin, C., and Nevatia, R., 1998. Building Detection and Description from a Single Intensity Image. *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 101-121.

Lowe, D.G., 2004. Distinctive Image Features from Scale Invariant Keypoints. *International Journal of Computer Vision*, Vol. 20, No. 2, pp. 91-110.

Madsen, C.B., and Christensen, H.I., 1995. Modeling and Testing the Stability of Edge Segments: Length and Orientation. In: Scandinavian Conference on Image Analysis, pp. 1011-1019.

Manual of Photogrammetry, 1980. Fourth Edition, American Society for Photogrammetry and Remote Sensing.

Manual of Photogrammetry, 2004. Fifth Edition, American Society for Photogrammetry and Remote Sensing.

Matlab, 2009. MATLAB 7.8.0, The MathWorks Inc., Natick, MA, 2009.

Meucci, A., 2005. Risk and Asset Allocation. Springer.

Mikhail, E.M., Bethel, J.S., and McGlone, J.C., 2001. Introduction to Modern Photogrammetry.

Mikolajczyk, K., and Schmid, C., 2005. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1615-1630.

Mohan, R., and Nevatia, R., 1989. Using Perceptual Organization to Extract 3-D Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 11, pp. 1121-1139.

Mohan, R., Medioni, G., and Nevatia R., 1989. Stereo Error Detection, Correction and Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11., No. 2, pp. 113-120.

Moons, T., Frère, D., Vandekerckhove, J., and Van Gool, L., 1998. Automatic Modelling and 3D Reconstruction of Urban House Roofs from High Resolution Aerial Imagery. In: Proc. 5th European Conference on Computer Vision, Freiburg, Germany, pp. 410-425.

Mundy, J.L., 1993. The Relationship between Photogrammetry and Computer Vision. In: International Society for Optical Engineering, San Jose California.

Nagabhushan, P., Guru, D.S., and Shekar, B.H., 2005. Eigen Transformation Based Edge Detector for Gray Images. *Lecture Notes in Computer Science 3776*, pp. 434-440.

Neumann, K.J., 2008. Trends for Digital Aerial Mapping Cameras. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII. Part B1, Beijing, pp. 551-554.

Nevatia, R., and Babu, K., 1980. Linear Feature Extraction and Description. *Computational Graphics Image Processing*, Vol. 13, 257-269.

Nitzberg, M., and Shiota, T., 1992. Nonlinear Image Filtering with Edge and Corner Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 8, pp. 826-833.

Noronha, S., and Nevatia, R., 2001. Detection and Modeling of Buildings from Multiple Aerial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 5, pp. 501-518.

Ó Conaire, C., O'Connor, N.E., and Smeaton, A.F., 2007. An Improved Spatiogram Similarity Measure for Robust Object Localisation. In: ICASSP, Hawaii.

Ohta, Y., and Kanade, T., 1985. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 7., No. 2, pp. 139-154.

Paragios, N., Chen, Y., and Faugeras, O., 2005. Handbook of Mathematical Models in Computer Vision. Springer.

Park, S., Lee, K., and Lee, S., 2000. A Line Feature Matching Technique Based on an Eigenvector Approach. *Computer Vision and Image Understanding*, Vol. 77, pp. 263-283.

Perona, P., and Malik, J., 1987. Scale-space and edge detection using anisotropic diffusion. In: Proceedings of IEEE Computer Society Workshop on Computer Vision, pp. 16-22.

Petrie, G., 2003. Airborne Digital Frame Cameras - The Technology is Really Improving. *GeoInformatics*, Vol. 6, No. 7, pp. 18-27.

Petrie, G., 2006. Airborne Digital Imaging Technologies. In: 6th International Conference on Laser Scanning and Digital Aerial Photography, Moscow, Russia.

Petrie, G., 2009. The IGI DigiCAM Range: Modular Cameras; Multiple Configurations. *GeoInformatics*, Vol. 12, No. 6, pp. 60-63.

Petrie, G., and Walker, A.S., 2007. Airborne Digital Imaging Technology: A New Overview. *Photogrammetric Record*, Vol. 22, No. 119, pp. 203-225.

Poli, D., 2005. Modeling of Spaceborne Linear Array Sensors. Ph.D. thesis, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zurich.

Princen, J., Illingworth, J., Kittler, J., 1990. A Hierarchical Approach to Line Extraction Based On the Hough Transform. *Computer Vision Graphics Image Processing*, Vol. 52, No. 1, pp. 57-77.

Ressl, C., 2004. Geometry, Constraints and Computation of the Trifocal Tensor. PhD Thesis, Vienna.

Roux, M., and McKeown, D. M., 1994. Feature Matching for Building Extraction from Multiple Views. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, Washington, pp. 46-53.

Rutzinger, M., Rottensteiner, F., and 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*, Vol. 2, No. 1, pp. 11-20.

Sahar, L., and Krupnik, A., 1999. Semiautomatic Extraction of Building Outlines from Large-scale Aerial Images. *Photogrammetric Engineering and Remote Sensing*, Vol. 65, No. 4, pp. 459-465.

Schenk, T., 1999. *Digital Photogrammetry, Volume I: Background, Fundamentals, Automatic Orientation Procedures*, TerraScience.

Schmid, C., and Zisserman, A., 1997. Automatic Line Matching Across Views. In: *Proceedings of Computer Vision and Pattern Recognition*, pp. 666–671.

Schmid, C., and Zisserman, A., 2000. The Geometry and Matching of Lines and Curves over Multiple Views. *International Journal of Computer Vision*, Vol. 40, No. 3, pp. 199-233.

Scholze, S., 2000. Exploiting Color for Edge Extraction and Line Segment Multiview Matching. Technical Report BIWI-TR-257, Communication Technology Laboratory, Computer Vision Group, ETH, Zurich, Switzerland.

Scholze, S., Moons, T., Ade, F., and Van Gool, L., 2000. Exploiting Color for Edge Extraction and Line Segment Stereo Matching. In: *International Archives of Photogrammetry and Remote Sensing*, 815–822.

Shao, J., Mohr, R., and Fraser C., 2000. Multi-Image Matching Using Segment Features. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXIII, Part B3. Amsterdam.

Shekar, B.H., Guru, D.S., and Nagabhushan, P., 2006. Object Recognition through the Principal Component Analysis of Spatial Relationship amongst Lines. *Lecture Notes in Computer Science* 3851, pp. 170-179.

Stassopoulou, A., and Caelli, T., 2000. Building Detection using Bayesian Networks. *International Journal of Pattern Recognition and Artificial Intelligence*. Vol. 14, No. 6, pp. 715-733.

Steger, C., 1998. An Unbiased Extractor of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20., No. 2, pp. 113-125.

Stilla, U., Michaelsen, E., 1997. Semantic Modelling of Man-made Objects by Production Nets. In: Gruen A, Baltsavias EP, Henricsson O (eds) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser, pp. 43-52.

Suveg, I., and Vosselman, G., 2004. Reconstruction of 3D Building Models from Aerial Images and Maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58, pp. 202-224.

Taillandier, F., and Deriche, R., 2002. Reconstruction of 3d Linear Primitives from Multiple Views for Urban Areas Modelisation. In: Proceedings Photogrammetric Computer Vision, (Vol B), Graz.

Tola E., Lepetit, V., and Fua, P., 2010. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 815-830.

Toth, C., 2009. The State-of-the Art in Airborne Data Collection Systems – Focused on LIDAR and Optical Imagery. Dieter Fritsch (Ed.), Photogrammetric Week 2009, pp. 147-161.

Turker, M., and San, B. T., 2004. Detection of collapsed buildings caused by the 1999 Izmit, Turkey earthquake through digital analysis of post-event aerial

photographs. *International Journal of Remote Sensing*, Vol. 25, No. 21, pp. 4701-4714.

USGS, 2010. Certification Report for the Microsoft Vexcel UltraCamD, UltraCamX, UltraCamXp, and UltraCamXp WA Models. Digital Aerial Sensor Type Certification.

Wang, Z., Wu, F., and Hu, Z., 2009. MSLD: A Robust Descriptor for Line Matching. *Pattern Recognition*, Vol. 42, pp. 941-953.

Web 1, 2011. Website of Institute of Photogrammetry, Stuttgart. Retrieved 01 August 2011, from the World Wide Web:  
[http://www.ifp.uni-stuttgart.de/forschung/photo/test\\_site/index.en.html](http://www.ifp.uni-stuttgart.de/forschung/photo/test_site/index.en.html)

Weickert, J., 1997. A Review of Nonlinear Diffusion Filtering, Scale Space Theory in Computer Vision, Lecture Notes in Computer Science, Vol. 1252, Springer, pp. 3-28.

Weijer, J., Gevers, T., and Bagdanov, A.D., 2006a. Boosting Color Saliency in Image Feature Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 1, pp. 150-156.

Weijer, J., Gevers, T., and Smeulders, A.W.M., 2006b. Robust Photometric Invariant Features from the Color Tensor. *IEEE Transactions on Image Processing*, Vol. 15, No. 1, pp. 118-127.

Werner, T., 2002. Matching of Line Segments across Multiple Views: Implementation Description. Visual Geometry Group, UK.

Werner, T., and Zisserman, A., 2002. New Techniques for Automated Architectural Reconstruction from Photographs. In: European Conference on Computer Vision. <http://cmp.felk.cvut.cz/cmp/software/lmatch/>

Wolf, P.R., and Dewitt, B.A., 2000. Elements of Photogrammetry with Applications in GIS, 3rd Edition.

Woo, D.M., Park, D.C., Han, S.S., and Beack, S., 2009. 2D Line Matching Using Geometric and Intensity Data. In: International Conference on Artificial Intelligence and Computational Intelligence.

Xiao, J., Gerke, M., and Vosselman, G., 2010. Automatic Detection of Buildings with Rectangular Flat Roofs from Multi-View Oblique Imagery. In: Proceedings of Photogrammetric Computer Vision and Image Analysis Conference, Paris, IntArchPhRS Vol. 38(3A), pp. 251-256.

Yip, R.K.K., and Ho, W.P., 1996. Multi-level Based Stereo Line Matching with Structural Information Using Dynamic Programming. In: International Conference on Image Processing, Vol. 2, pp. 341-344.

You, H., and Shiqiang, Z., 2006. 3D Building Reconstruction from Aerial CCD Image and Sparse Laser Sample Data. *Optic and Lasers in Engineering*, Vol. 44, pp. 555-566.

Zhang, C., and Baltsavias, E. P., 2000. Edge Matching And 3D Road Reconstruction Using Knowledge-Based Methods. In: Schriftenreihe der Fachrichtung Geodaesie, Darmstadt, Germany, 10, pp. 251-265.

Zhang, L., 2005. Automatic Digital Surface Model (DSM) Generation from Linear Array Images. PhD Thesis, Institute of Geodesy and Photogrammetry, Swiss Institute of Technology, Zurich.

Zhongliang, F., and Zhiqun, S., 2008. An Algorithm of Straight Line Features Matching on Aerial Imagery. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B3b. Beijing.

## APPENDIX A

The details of the definitions and various forms of mathematical models can be found in elsewhere (Manual of Photogrammetry, 1980; Koch, 1999; Manual of Photogrammetry, 2004; Förstner, 2005). Here, only the derivation of the well-known generic Gauss-Helmert estimation model with constraints is provided. The Gauss-Markov model (with constraints) which can also be expressed as a special type of the Gauss-Helmert model is also stated. The following part is based on the least squares estimation section provided in Manual of Photogrammetry (2004).

The Gauss-Helmert model with constraints is expressed by a functional model composed of not only condition equations  $g(l + \hat{v}, \hat{p}) = 0$ , but also a set of constraints  $h(\hat{p}) = 0$ . In those equations,  $l$ ,  $p$ , and  $v$  denote the observations, unknown parameters, and corrections, respectively. The operator (^) indicate the estimated forms of the related term. Since the given functional model can be non-linear, the linearized form of the Gauss-Helmert model with constraints can be written as:

$$A\widehat{\Delta p} + B^t \hat{v} = c_g \quad (\text{A.1})$$

$$H^t \widehat{\Delta p} = c_h \quad (\text{A.2})$$

with

$$c_g = -g(\hat{l}^{(0)}, \hat{p}^{(0)}) - B^t(l - \hat{l}^{(0)}) \quad (\text{A.3})$$

$$c_h = -h(\hat{p}^{(0)}) \quad (\text{A.4})$$

$$\widehat{\Delta p} = p - \hat{p}^{(0)} \quad (\text{A.5})$$

In Eq. A.1 and A.2, the matrices  $A$  and  $H$  contain the partial derivatives with respect to parameters, whereas the matrix  $B$  involves the partial derivatives with

respect to observations. The terms  $c_g$  and  $c_h$  (contradiction vectors) contain the differences between the function values  $g$  and  $h$ , respectively, evaluated at the approximate values  $\hat{p}^{(0)}$  and their ideal values is 0. Note also that the contradiction vector  $c_g$  depends on the observations  $l$  and is iteratively refined.

Assuming that an initial covariance matrix  $C_{ll}^{(0)}$  of the observations  $l$  is known and related to the true covariance matrix  $C_{ll}$  by  $C_{ll} = \sigma_0^2 C_{ll}^{(0)}$  with a possibly unknown variance factor  $\sigma_0^2$ . In a weighted least square sense, it is necessary to minimize a quadratic equation that involves the square of the residuals and inverse of the initial covariance matrix  $C_{ll}^{(0)}$  :

$$\Omega = \hat{v}^t C_{ll}^{-1} \hat{v} = (\hat{l} - l)^t C_{ll}^{-1} (\hat{l} - l) \quad (\text{A.6})$$

under the linear constraints  $A\widehat{\Delta p} + B^t \hat{v} = c_g$  and  $H^t \widehat{\Delta p} = c_h$ . To do that, with the help of Lagrange multipliers  $(\lambda, \mu)$ , it is necessary to minimize the form:

$$\Phi = (\hat{l} - l)^t C_{ll}^{-1} (\hat{l} - l) + 2\lambda^t (A\widehat{\Delta p} + B^t \hat{v} - c_g) + 2\mu^t (H^t \widehat{\Delta p} - c_h) \quad (\text{A.7})$$

After setting the partial derivatives of  $\Phi$  to zero, we obtain with  $\hat{v} = \hat{l} - l$

$$\frac{1}{2} \left( \frac{\partial \Phi}{\partial \hat{l}} \right)^t = C_{ll}^{-1} \hat{v} + B\lambda = 0 \quad (\text{A.8})$$

$$\frac{1}{2} \left( \frac{\partial \Phi}{\partial \hat{p}} \right)^t = A^t \lambda + H\mu = 0 \quad (\text{A.9})$$

$$\frac{1}{2} \left( \frac{\partial \Phi}{\partial \lambda} \right)^t = -c_g + A\widehat{\Delta p} + B^t \hat{v} = 0 \quad (\text{A.10})$$

$$\frac{1}{2} \left( \frac{\partial \Phi}{\partial \mu} \right)^t = -c_h + H^t \widehat{\Delta p} = 0 \quad (\text{A.11})$$

From Eq. A.8, follows the relation,

$$\hat{v} = -C_{ll} B \lambda \quad (\text{A.12})$$

Substituting Eq. A.12 into Eq. A.10 yields

$$\lambda = (B^t C_{ll} B)^{-1} (A \widehat{\Delta p} - c_g) \quad (\text{A.13})$$

Thereafter, substituting Eq. A.13 into Eq. A.9 yields the normal equation system:

$$\begin{bmatrix} A^t (B^t C_{ll} B)^{-1} A & H \\ H^t & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta p} \\ \mu \end{bmatrix} = \begin{bmatrix} A^t (B^t C_{ll} B)^{-1} c_g \\ c_h \end{bmatrix} \quad (\text{A.14})$$

With the solution of the normal equation system, the Lagrange multiplier  $\lambda$  can be obtained from Eq. A.13, which then yields the estimated residuals in Eq. A.12. The estimated variance factor is given by

$$\hat{\sigma}_0^2 = \frac{\hat{v}^t C_{ll}^{-1} \hat{v}}{R} \quad (\text{A.15})$$

with the redundancy term  $R = G + H - U$ , where  $G$ ,  $H$  and  $U$  denote the number of equations for the observations, constraints and unknowns, respectively. The estimated covariance matrix of the estimated parameters can be obtained by  $\hat{C}_{\hat{p}\hat{p}} = \hat{\sigma}_0^2 C_{\hat{p}\hat{p}}$  where  $C_{\hat{p}\hat{p}}$  results from the inverted normal equation matrix:

$$\begin{bmatrix} C_{\hat{p}\hat{p}} & S \\ S^t & T \end{bmatrix} = \begin{bmatrix} A^t (B^t C_{ll} B)^{-1} A & H \\ H^t & 0 \end{bmatrix}^{-1} \quad (\text{A.16})$$

Different models derived for the evaluation of the results based on the estimated residuals  $\hat{v}$  and their covariance matrix  $C_{\hat{p}\hat{p}}$  can also be found in Manual of Photogrammetry (2004). Since the presented estimation model Gauss-Helmert with constraints is the most generic estimation model, the Gauss-Markov model with constraints can be obtained by replacing the matrix  $B$  with  $-I$ , and the well-known Gauss-Markov model can be obtained by replacing the matrix  $B$  with matrix  $-I$  and leaving out the constraints.

## CURRICULUM VITAE

### PERSONAL INFORMATION

Surname, Name: Ok, Ali Özgün  
Nationality: Turkish (T.C.)  
Date and Place of Birth: 28 August 1980, Ankara  
Marital Status: Married  
Phone: +90 505 367 3440  
e-mail: ozguneo@gmail.com

### EDUCATION

Degree	Institution	Graduation Year
M.Sci.	Geodetic and Geographic Information Technologies Dept.	2005
B.Sci.	Selcuk Univ., Civil Engineering Dept.	2001
High School	Private Yüce Science School	1997

### WORK EXPERIENCE

Year	Place	Enrollment
2004 – Present	METU, Geodetic and Geographic Information Technologies Dept.	Research Assistant

### PUBLICATIONS

1. Ok, A.O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., Toprak, V., 2011. Accurate Matching and Reconstruction of Line Features from Ultra High Resolution Stereo Aerial Images. In: Proceedings of ISPRS Hannover Workshop, High Resolution Earth Imaging for Geospatial Information, Hannover, Germany.
2. Ok, A.Ö., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., Toprak, V., 2011. Stereo Hava Fotoğraflarından Doğrusal Çizgilerin Otomatik Eşlenmesi ve Geri Çatımı için Çift-Tabanlı Yeni Bir Yaklaşım. In: Türkiye Ulusal

Fotogrametri ve Uzaktan Algılama Birliđi VI. Teknik Sempozyumu, 23-26 Şubat, Antalya, Türkiye. (In Turkish)

3. Ok, A.O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., Toprak, V., 2010. A New Straight Line Reconstruction Methodology from Multi-Spectral Stereo Aerial Images. In: Proceedings of Photogrammetric Computer Vision and Image Analysis Conference, Paris, IntArchPhRS Vol. 38(3A), pp. 25-30.
4. Wegner, J.D., Ok, A.O., Thiele, A., Rottensteiner, F., Soergel, U., 2010. Urban Building Detection from Optical and InSAR Features Exploiting Context. In: Proceedings of Photogrammetric Computer Vision and Image Analysis Conference, Paris, IntArchPhRS Vol. 38(3A), pp. 239-244.
5. Ok, A.O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., Toprak, V., 2010. A Stereo Line Matching Technique For Aerial Images Based On A Pair-Wise Relation Approach. In: Proceedings of ISPRS Istanbul Workshop, Modeling Of Optical Airborne and Spaceborne Sensors, Istanbul, Turkey.
6. Ok, A.Ö., Wegner, J.D., Heipke, C., Rottensteiner, F., Sörgel, U., Toprak, V., 2010. Çok Bantlı Stereo Hava Fotoğraflarından Doğrusal Çizgilerin Otomatik Geri-Çatımı İçin Yeni Bir Yaklaşım. In: Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Gebze, Türkiye. (In Turkish)
7. Ok, A.O., 2009. Automated Description of 2-D Building Boundaries from A Single Color Aerial Ortho-Image. In: Proceedings of ISPRS Hannover Workshop, High Resolution Earth Imaging for Geospatial Information, Hannover, Germany.
8. Ok, A.O., 2008. Robust Detection of Buildings from A Single Color Aerial Image. In: Proceedings of GEOBIA 2008 - Pixels, Objects, and Intelligence, Calgary, Alberta, Canada.
9. Ok, A.Ö., Güllüođlu, C., 2008. Temel Düzey Pankromatik Quickbird Görüntüsünün Farklı Sensör Modelleri Kullanılarak Orto-Rektifikasyonu. In: Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Kayseri, Türkiye. (In Turkish)
10. Özdarıcı, A., Ok, A.Ö., 2007. Mean-Shift Segmentasyon Algoritması Destekli 'En Büyük Olasılık' Sınıflandırma Yöntemi Kullanılarak Tarım Alanlarının Sınıflandırılması. In: Ulusal Coğrafi Bilgi Sistemleri Kongresi, Trabzon, Türkiye. (In Turkish)
11. Ok, A.Ö., Türker, M., 2007. Çok Bantlı Bilsat-1 Görüntülerinin Self-Kalibrasyonu ve Ortorektifikasyonu. In: Ulusal Coğrafi Bilgi Sistemleri Kongresi, Trabzon, Türkiye. (In Turkish)

12. Ok, A.O., 2007. Semi-Automated Registration of BILSAT-1 XS Imagery via Global Orthorectified Landsat ETM+ Data Set. In: Proceedings of the 3rd International Conference on Recent Advances in Space Technologies, Istanbul, Turkey.
13. Ok, A.O., Turker, M., 2007. Geometric Calibration and Orthorectification of BILSAT-1 Imagery. In: Proceedings of American Society for Photogrammetry and Remote Sensing Annual Conference, Tampa, Florida, USA.
14. Ok, A.O., Turker, M., 2006. Orthorectification of Bilsat Imagery using Rigorous and Simple Geometric Models. In: International Archives of Photogrammetry and Remote Sensing, Band XXXVI 1/ W41. Ankara, Turkey.
15. Ok, A.O., Turker, M., 2006. Comparison of Different Mathematical Models on the Accuracy of the Orthorectification of ASTER Imagery. In: International Archives of Photogrammetry and Remote Sensing, Band XXXVI 1/ W41. Ankara, Turkey.
16. Ok, A.O., Turker, M., 2005. DEM Generation and Accuracy Assessment from Stereo ASTER Imagery. In: Proceedings of the 2nd International Conference on Recent Advances in Space Technologies, Istanbul, Turkey, pp. 523-528.
17. Ok, A.Ö., Türker, M., 2005. Stero ASTER Uydu Görüntülerinden Sayısal Yükseklik Modeli Oluşturma ve Doğruluk Analizleri. In: Ege Coğrafi Bilgi Sistemleri Sempozyumu, İzmir, Türkiye. (In Turkish)
18. Ok, A.O., 2005. Accuracy Assessment of the DEM and Orthoimage Generated from ASTER. M.Sci. Thesis, Middle East Technical University, Ankara, Turkey.