

Image Registration Under Arbitrarily-Shaped Local Illumination Variations

by

Mohamed Mahmoud Aly Fouad Abdelmotagally, M. Sc.

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering (OCIECE)

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada, K1S 5B6

July 2010

©Copyright 2010, Mohamed Mahmoud Aly Fouad Abdelmotagally



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-70534-6
Our file *Notre référence*
ISBN: 978-0-494-70534-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

Image registration is the process of geometrically matching the corresponding pixels in images captured for the same scene at different times and/or sensors from various perspectives. Image registration approaches can be classified according to many viewpoints, such as various transformation models, spatial versus frequency domain, single- versus multi-modal, and intensity-based versus feature-based. Image registration plays a central role in several applications, such as remote sensing, medical imaging, computer vision, change detection, and super-resolution.

The problem is that the geometric registration precision could be impacted due to local illumination variations. Thus, any subsequent processing would be easily negatively affected. In early research, the registration process assumed brightness constancy. Recently, some research has incorporated illumination variations in the registration process in a limited manner, such as using a global or an affine illumination model.

In this dissertation, we address the impacts of the arbitrarily-shaped locally variant illuminations on the geometric registration precision. Given a perfect camera following a pin-hole camera model and no object motions exist in the scene, we propose an intensity-based image registration model that can handle arbitrarily-shaped local illumination variations, ASLIV. Then, the ASLIV model

is cast in a registration approach whose idea is based on iteratively segmenting the absolute image difference between the input images into distinct illumination regions. Assuming gain and offset uniformity along with each region, the proposed approach applies different M -estimators simultaneously. Each estimator has its own objective function that is assigned to a certain illumination region. The residuals of an illumination region are then differently penalized in accordance with their own objective function to minimize the registration error. In addition, some areas located on the boundary of each region could be mis-segmented due to the iterative process of creating the illumination regions. To lessen the negative impacts of such misclassified areas, a weighting function has been used. The proposed approach is cast in an iterative coarse-to-fine scheme to allow for large motions.

Experiments show that the proposed approach yields clear improvements in terms of geometric registration precision and illumination correction with a slight increase in computational time compared to competing approaches. As well, the proposed approach shows more resistance against segmentation perturbations as opposed to others. Real and simulated image pairs are employed in the experiments. The performance of competing approaches is evaluated using: normalized cross-correlation (NCC), structural similarity (SSIM) index, and peak signal-to-noise ratio (PSNR).

I would like to dedicate this work to my mother, my father, my wife, my sister
and my two princesses: Nada and Malak.

Acknowledgments

Thanks GOD.

First and foremost, I should like to express my deepest thanks to my supervisor Prof. Richard M. Dansereau for his guidance, support, and friendship. He is my role model of an exceptional scientist and teacher. He has always been supportive, not only when I made progress but also when I make mistakes. I have also learned from him how to be professional in an academic career. It is my great privilege to study my Ph.D. research under his supervision.

I am also grateful to Prof. Anthony D. Whitehead for generously sharing his intelligent ideas and comments that are constant sources of encouragement. He is also appreciated for his friendship, empathy, and great sense of humor.

Special thanks goes to Mrs. Darlene Hebert and Mrs. Coleen Kornelsen, who provided me with constant help and assistance since the first day.

Table of Contents

Abstract	iii
Acknowledgments	vi
Table of Contents	vii
List of Figures	xi
List of Tables	xiii
Nomenclature	xv
1 Introduction	1
1.1 Motivation & Problem Statement	1
1.2 Problem Challenges	3
1.3 Dissertation Outline	5
1.4 Dissertation Contributions	8
1.5 Dissertation Organization	9
2 Literature Review	11
2.1 Introduction	11
2.2 Feature-based Methods	12

2.2.1	Iterative Closest Point Method	13
2.2.2	EM-Iterative Closest Point (ICP) Method	14
2.2.3	Keypoint Matching With Invariant Descriptors	15
2.3	Intensity-based Methods	17
2.3.1	Fourier-based Registration Methods	17
2.3.2	Correlation-based Registration Methods	19
2.4	Motion Field Representation	21
2.4.1	Parametric and Nonparametric Motion Models	21
2.4.2	Common Motion Models	23
2.4.3	Regions of Support for Motion Models	23
2.5	Relating Motion and Image Intensities	25
2.6	Applications	28
2.7	Summary	30
3	Proposed Registration Model Under Local Illumination Variations	32
3.1	Introduction	32
3.2	Image Formation Model	33
3.3	The Proposed Model	34
3.4	Image Registration using LSE	40
3.5	Data Set Description	46
3.6	Performance Evaluation	48
3.7	Analysis on Stopping Conditions	50
3.8	Summary on the LS-ASLIV Approach	52
3.9	Experiments and Results	53
3.9.1	Experiments on Simulated Data Sets	53
3.9.2	Experiments on the Real Data Set	57

3.10	Analysis & Discussion	59
4	Image Registration Using The M-estimator	65
4.1	Introduction	65
4.2	A General Framework of M -estimation With ASLIV	66
4.3	An Iterative Scheme Using the M -estimation With the ASLIV Model	67
4.4	Summary on the HM-ASLIV Approach	72
4.5	Experiments & Results	74
4.5.1	Experiments on Simulated Data Sets	74
4.5.2	Experiments on the Real Data Set	79
4.6	Analysis & Discussion	82
5	Image Registration Using Weighted Region-based M-estimators	84
5.1	Introduction	84
5.2	Weighted Region-Based M -estimators For Image Registration . . .	86
5.2.1	A General Framework of M -estimators With Region-based ASLIV Model	87
5.2.2	A Weighting Function For Segmentation Deviations	87
5.3	Proposed Approach Using M -estimator & Selected Functions	90
5.3.1	M -estimation Using Huber & Tukey Functions	91
5.4	Implementation Setup	95
5.5	Summary on The Proposed WH-ASLIV Approach	97
5.6	Experiments & Results	99
5.6.1	Experiments on Simulated Data Sets	99
5.6.2	Experiments on the Real Data Set	102
5.6.3	Experiments on Simulated Segmentation Deviations	105
5.7	Analysis & Discussion	110

6 Conclusions & Contributions	113
6.1 Conclusions	114
6.2 Contributions	115
6.3 Publications	116
6.4 Future Work	117
List of References	119
Appendix A Real Image Pairs	130
Appendix B Segmentation Deviations Results	135

List of Figures

2.1	Natural canyon and LANDSAT pairs captured at different times. . .	31
3.1	With <i>a priori</i> known information, the figure shows two different sets of distinct illumination levels of two input images, yielding certain arbitrarily-shaped regions, each has its own gain and offset.	38
3.2	Binary masks of distinct illumination regions at two different iterations.	39
3.3	An illustrative scheme of the proposed registration approach LS-ASLIV using the <i>k</i> -means algorithm for segmentation	47
3.4	Examples on “ <i>J</i> =3” and “ <i>J</i> =4” data sets.	49
3.5	Normalized AIDs of the aligned images for a “ <i>J</i> =4” simulated pair using competing approaches.	62
3.6	Image mosaics of the natural pair in Fig. 2.1(a,b).	63
3.7	Gamma-corrected AIDs of the real pair #B, $\gamma=1.7$	64
4.1	Huber function and its influence function.	69
4.2	Normalized AIDs of a “ <i>J</i> =3” simulated pair (#34).	78
4.3	Gamma-corrected AIDs of for the real pair in Fig. 2.1(c,d) using competing approaches.	80
4.4	Blow up views of the white boxes in Fig. 4.3.	81

5.1	Absolute image deference is segmented into small and large residuals.	86
5.2	Blow up view of the boundary layers of two adjacent regions.	90
5.3	Normalized AIDs of the resulting aligned pairs for a “ $J=3$ ” simulated pair.	103
5.4	Gamma-corrected AIDs of for the real pair in Fig. 2.1(c,d) using competing approaches.	106
5.5	Blow up views of the white boxes of Fig. 5.4.	107
5.6	The absolute error (AE) of six affine parameters using competing approaches at different segmentation noise levels.	109
A.1	Real image pair #C.	131
A.2	Real image pair #D.	131
A.3	Real image pair #E.	132
A.4	Real image pair #F.	132
A.5	Real image pair #G.	133
A.6	Real image pair #H.	133
A.7	Real image pair #I.	134
A.8	Real image pair #J.	134

List of Tables

2.1	The hierarchy of 2D parametric motion models [1].	23
3.1	The average of absolute errors of estimated affine parameters using computing approaches with the “ $J=4$ ” data set.	56
3.2	Image quality measures express the correlation between the overlapping areas of the two registered images of the “ $J=4$ ” data set using competing approaches.	57
3.3	Computational time for a random subset of the “ $J=4$ ” data set using competing approaches.	58
3.4	NCC for the real data set using competing approaches.	59
4.1	Examples of potential functions $\rho(\cdot)$ that satisfy certain properties. .	68
4.2	The AAEs of the estimated affine parameters of the “ $J=3$ ” data set. .	75
4.3	The AAEs of the estimated affine parameters of the “ $J=4$ ” data set. .	75
4.4	SSIM indexes and PSNR values for a random subset of the “ $J=3$ ” data set using competing approaches.	77
4.5	Computational time for a random subset of the “ $J=3$ ” data set using competing approaches.	77
4.6	SSIM indexes and PSNR values for a random subset of the “ $J=4$ ” data set using competing approaches.	77

4.7	Computational time for a random subset of the “ $J=4$ ” data set using competing approaches.	79
4.8	NCC values for the real data set using competing approaches.	82
5.1	The AAEs of the estimated affine parameters of the “ $J=3$ ” data set.	101
5.2	The AAEs of the estimated affine parameters of the “ $J=4$ ” data set.	101
5.3	SSIM indexes for a random subset of the “ $J=3$ ” data set.	101
5.4	PSNR values for a random subset of the “ $J=3$ ” data set.	102
5.5	Computational time for a random subset of the “ $J=3$ ” data set.	102
5.6	SSIM indexes for a random subset of the “ $J=4$ ” data set.	102
5.7	PSNR values for a random subset of the “ $J=4$ ” data set.	104
5.8	Computational time for a random subset of the “ $J=4$ ” data set.	104
5.9	NCC values for the real data set using competing approaches.	105
B.1	Percentage increase of AAEs of the estimated parameter a_1	136
B.2	Percentage increase of AAEs of the estimated parameter a_2	136
B.3	Percentage increase of AAEs of the estimated parameter a_3	137
B.4	Percentage increase of AAEs of the estimated parameter a_4	137
B.5	Percentage increase of AAEs of the estimated parameter a_5	138
B.6	Percentage increase of AAEs of the estimated parameter a_6	138

Nomenclature

AAE	average of absolute error
AID	absolute image difference
AIM	affine illumination model
ASLIV	arbitrarily-shaped local illumination variations
EM	Expectation Maximization
GDA	gradient-descent algorithm
GIM	global illumination model
GIR	Geometric image registration
GNA	Gauss-Newton algorithm
GRP	geometric registration precision
ICP	Iterative Closest Point
LMA	Levenberg-Marquardt algorithm
LS	least-squares
LSE	least-squares estimator
MSE	mean square error
NCC	normalized cross-correlation
PSNR	peak signal-to-noise ratio
SR	super-resolution
SSD	sum of squared difference
SSIM	structure similarity

Chapter 1

Introduction

In this chapter, we state the research motivations and problem description in Section 1.1. As well, possible problem challenges are fully presented in Section 1.2. Then, proposed solutions to those challenges are outlined in Section 1.3. The main contributions of this dissertation are clearly given in Section 1.4. Finally, dissertation structure is provided in Section 1.5.

1.1 Motivation & Problem Statement

Geometric image registration (GIR) is the process of matching the corresponding pixels in images captured for the same scene with variant sensors and exposures at different times. Image capturing time plays an effective factor in the registration step. Hence, there may exist global and/or local motions as well as occlusions in the captured images. GIR is a common pre-processing step in

many applications [2], such as generating panoramic image mosaics [3], super-resolution (SR) enhancement [4], image stitching [5], remote-sensing applications [6], change detection [7], and medical imaging [8]. For all of these applications, sub-pixel precision in registration is necessary for satisfactory post-processing results. Unfortunately, local illumination variations between some image sets tend to degrade or even spoil the resulting geometric registration.

To clearly comprehend the problem, let us illustrate with the following real example. Consider two images for the same scene captured at two different times of day, *e.g.*, one in the morning and one at noon, and are taken with the same exposure and a perfect camera following a pin-hole camera model. As well, we assume no object motions exist. In presence of a light source, such as our sun, the image pair will contain multiple areas with different illumination levels due to the structure of the objects in the scene. Within this image pair, one can observe that local illumination regions with arbitrary shapes exist which leads to significant degradation in registration performance using conventional models. The scene is thus assumed to be static from an object point of view, but not static in terms of locally variant illuminations. Hence, the principle objective of this dissertation is to *improve the geometric registration precision (GRP)* by minimizing the factors affecting this process, such as illumination variations.

Existing image registration approaches can be classified into two broad categories: feature-based and intensity-based. The former attempts to extract a number of corresponding landmarks or features from image pairs to be registered, which are used to estimate the geometric transformation. The major problem of feature-based approaches is their high sensitivity to the feature extraction and matching process, especially in the presence of local illumination changes. As well, accurate sub-pixel results are hard to obtain because of the sparse nature of

the feature set [9].

Intensity-based approaches aim to find a transformation function, such that the intensity differences between corresponding pixels are minimized. Although most of the image registration approaches assume *brightness constancy*, recently some researchers have incorporated illumination variations in their registration models. What motivates us to start this research is that current approaches, however, tend to account for illumination variations in a global [10–12] or affine fashion [13]. Local illumination changes with arbitrary shapes, which can take place in real imaging scenarios, are not fully considered. Thereby, any highly demanding applications for accurate sub-pixel geometric registration, *e.g.*, SR, would be easily impacted.

1.2 Problem Challenges

Before we proceed in exhibiting the problem challenges, first we should differentiate between photometric corruption and illumination variations from an image registration perspective. Photometric corruption takes place due to issues that relate to the camera itself, such as exposure time, camera response function, or sensor noise. Illumination variations are concerned with the scene itself, such as capturing two images for the same scene at two different times of day, or year, resulting in multiple areas with distinct illumination levels. In this dissertation, we focus on illumination variations only.

Photometric correction-based image registration approaches can be classified into three categories. First, images are photometrically registered using an

algorithm that is not vulnerable to geometric misalignments, followed by geometric registration [14]. Second, images are geometrically registered using an algorithm that is exposure-insensitive, followed by photometric registration [15]. Third, geometric registration and photometric correction are simultaneously performed [13]. The first two categories are limited to either assuming that there is no saturation or significant geometric misalignment, or knowing the camera response function in advance leading to accumulated errors. Similarly, we take into account locally variant illuminations instead of photometric corruption using the third category.

Current intensity-based registration approaches do not fully consider the local illumination changes. In [10–12], a global illumination model is proposed in which each input image is assumed to have gain and offset scalars. In [13], an affine illumination model is proposed with triangular or quadrilateral region support. As well, it can be noticed that arbitrariness of shapes of those variations is typically assumed to be predefined. These assumptions cannot hold for real imaging scenarios. Thus, we intend to model arbitrary-shaped local illumination variations, ASLIV. Since we believe that intensity-based approaches avoid some of the pitfalls of the feature-based approaches for high precision registration, the proposed registration model will be cast in an intensity-based approach.

The first challenge is to formulate an intensity-based registration model to cope with images having locally variant illuminations, in order to ultimately achieve improving the GRP jointly with illumination correction. These local illumination changes compose a set of illumination regions. With *a priori* information on illumination regions, there would be no problem. However, *a priori* illumination information is rarely available for real imaging scenarios. A second challenge appears when there is no *a priori* information on how the illumination

regions are segmented.

A third challenge arises when choosing an appropriate estimator for the minimization problem. A commonly used estimator is the least-squares estimator (LSE). Such an estimator can be used for its simplicity in mathematical derivations and implementation. However, it squares the residuals. The behavior of the quadratic cost function in LSE could negatively affect the GRP.

Other estimators are robust than LSE, such as M -estimator [16, 17]. The latter can generally employ any objective function with certain aspects. A fourth challenge arises when choosing such a cost function that should follow specific conditions. The most dominant aspect of the objective function is a tuning parameter by which the registration residuals are differently penalized.

In case of large geometric deformations, a fifth challenge would arise. Given no object motions, geometric deformations, such as translation, rotation, scaling and skew are assumed in the input images.

Moreover, segmentation errors would occur during the creation of the illumination regions. The sixth challenge then arises during the registration process, when the boundary of each resulting illumination region contains some areas that belong to its adjacent region.

1.3 Dissertation Outline

In this section, we summarize how previous challenges are addressed and analyzed in the dissertation. First, we propose an intensity-based image registration model that can handle arbitrarily-shaped local illumination variations. We employ the ASLIV model with a LS estimator to accomplish sub-pixel geometric registration jointly with illumination correction as shown in [18]; this approach is

referred to LS-ASLIV. We noticed that the LS-ASLIV approach improves the GRP and the illumination of some regions compared to the global illumination approach shown in [10–12] and the affine illumination approach in [13]. However, one disadvantage of the LS-ASLIV is its sensitivity to outliers. Note that outliers in intensity-based registration approaches refers to the corresponding pixels that have large absolute intensity difference due to many factors, such as illumination variations in our case.

We replace the LSE with a robust M -estimator [16] that is not as sensitive to outliers. Many objective functions having certain characteristics can be employed with the M -estimator. The idea behind using a certain objective function is that the latter should have a tuning parameter by which the large and small residuals are differently penalized. We use the Huber function [16] because it has a threshold by which linear and quadratic penalties are assigned to large and small residuals, respectively. This approach, referred to as HM-ASLIV, yields improving GRP compared to the LS-ASLIV approach as shown in [19, 20]. However, the scalar threshold of any candidate cost function is not necessarily appropriate for all illumination regions.

Since the proposed model, ASLIV, assumes varying illumination regions, we propose a registration approach in which different M -estimators are simultaneously applied. Each employs a distinct objective function, with its own threshold being assigned to a certain illumination region whose residuals are segmented into large and small areas. Thus, we extend the ordinary formula of the Huber function from one threshold to a region-based expression with multiple thresholds. This approach can be referred as region-based M -estimators, RH-ASLIV. This latter approach improves the GRP and obtains illumination correction compared to the HM-ASLIV approach.

Without *a priori* known information, the illumination regions are iteratively created by segmenting the image similarities of the input images. There are many approaches to perform the segmentation step, however a clustering algorithm is used. Since, the proposed approach is cast in an iterative coarse-to-fine scheme to allow for large motions, another challenge is due to segmentation errors. Some areas located on the boundary of each region could belong to an adjacent region, thus being mis-segmented. Therefore, the proposed approach uses an uncertainty function to lessen the negative impacts of such misclassified areas. The proposed approach can be referred as weighted region-based M -estimators, WH-ASLIV. It addresses the aforementioned problems and overcomes the drawbacks of competing approaches.

Given no object motions, geometric deformations, such as translation, rotation, scaling and skew are assumed in the input images. To handle the large geometric deformations, we cast the minimization problem within an iterative coarse-to-fine scheme [21]. We'd like to stress that the coarse-to-fine scheme is also used to increase the geometric registration accuracy (GRA) by improving the chance of convergence, but that the focus of this dissertation is on improving sub-pixel precision on cases where convergence indeed occurs. As such, we do assume for image pairs tested that correct convergence does occur with approaches like GIM, AIM, or some other image registration approach, or that we have *a priori* knowledge that allows for a reasonable initial image alignment. From these initial rough registrations, the proposed approach will seek to improve the precision of the registration.

The registration process is initialized by a certain initial guess at the coarsest resolution level and updated by a determined amount for a definite number of

iterations per resolution level. The approach turns to the next finer level, once either the customized cost function is updated by less than a predefined threshold, or a maximum number of iterations has been reached. As well, some parameters are heuristically chosen, such as: the initial guess of the unknown parameters, the number of resolution levels, predefined threshold and maximum number of iterations.

1.4 Dissertation Contributions

In this section, we can summarize the main contributions of this dissertation:

1. Robustness against local illumination changes:

We proposed an intensity-based model to obtain a more precise sub-pixel geometric registration jointly with illumination correction. This model is robust against locally variant illumination regions in the images and outperforms the global and affine illumination models in terms of geometric registration precision as shown in [19]. A special case, when considering only two illumination regions, is published in [18].

2. Capability of differently penalizing local residuals:

The proposed approach simultaneously employs different robust M -estimators. We extend the ordinary formula of an objective function having single-threshold to be a spatially adaptive region-based function with multiple-threshold. Thus, each threshold is iteratively assigned to an illumination region. A framework of spatially adaptive M -estimators is achieved using a cost function with certain properties. A special case, when using the ordinary Huber function as cost function, is published in [20].

3. Robustness against segmentation errors:

The proposed approach exhibits more resistance versus mis-segmented areas located on the boundary layer of each illumination region due to segmentation. The proposed approach is able to reduce the negative effects of those mis-segmented areas by lessening their contributions throughout the registration process using a weighting function. Thus, geometric registration precision is further improved as well as illumination correction is obtained. This contribution and its precedent are submitted to [22].

1.5 Dissertation Organization

The rest of this dissertation is structured in five chapters as follows:

Chapter 2 gives a background on feature-based, and intensity-based image registration approaches with advantages and disadvantages for each.

Chapter 3 introduces an intensity-based image registration model that accounts for arbitrary-shaped local illumination variations; ASLIV. Then, a registration approach exploits the proposed model jointly with the LSE. The approach is cast in an iterative coarse-to-fine scheme. The real and simulated data sets are, then, fully described with commonly used performance evaluation measures. A heuristic analysis of some critical parameters and a discussion on implementation issues are also provided. Such an LSE approach results in improving in GRP compared to the global illumination approach. An analysis of the LS-ASLIV-based approach is introduced.

Chapter 4 contributes an approach to overcome the sensitivity of the LS-ASLIV-based approach using a more robust estimator, such as an M -estimator. The M -estimation approach is generally formulated using the proposed ASLIV

model with an objective function that should satisfy certain conditions, such as a tuning parameter by which the residuals can be differently penalized. The approach is cast in an iterative multi-resolution framework, yielding more precise estimates than those of the LS-ASLIV-based approach. This approach is analyzed and other related issues are discussed.

Chapter 5 proposes a new spatially adaptive registration approach that simultaneously addresses three problems: vulnerability to outliers, applying a unique threshold to the whole residuals' space and negative impacts of the mis-segmented areas. First, the proposed approach is generally presented using the proposed ASLIV model within a framework of different M -estimators. Second, each estimator should employ a distinct cost function, with its tuning parameters that are assigned to a certain illumination region. In turn, each region-based tuning parameter is iteratively adapted yielding two subregions of small and large residuals. Third, the proposed approach can use any uncertainty function that assigns designated weights to the pixels located on the boundary of each illumination region. Then, the proposed approach is also cast in an iterative coarse-to-fine scheme, producing improvements in terms of GRP and illumination correction compared to competing approaches. As well, the chapter explains some experiments performed to show the effectiveness of the proposed approach versus segmentation errors as opposed to other approaches. Discussion on limitations of the proposed approach is also provided.

Chapter 6 provides: research conclusions, publications resulting from this dissertation and possible future work directions.

Chapter 2

Literature Review

2.1 Introduction

Image registration is the process of geometrically matching corresponding sub-pixels in two images (or more) captured for the same scene at different times and/or by different sensors. The registration geometrically aligns two images (the reference and sensed images). A survey of image registration methods is given by Zitová and Flusser [2]. Both advantages and drawbacks of the reviewed methods are mentioned in this chapter as well as the applications relying on image registration. The main goal of this chapter is to provide a background of different methods of image registration.

This chapter is organized as follows. The reviewed methods are classified into two main categories: feature-based and intensity-based methods presented in Section 2.2 and Section 2.3, respectively. Section 2.4 discusses commonly used motion field representations that relates to the intensity-based methods in Section 2.5. Many applications requiring image registration as a common pre-processing step for satisfactory post-processing results are reviewed in Section 2.6. Finally, this chapter is concluded in Section 2.7.

2.2 Feature-based Methods

Feature-based methods start by extracting image geometric primitives, namely features, such as corners, curves, surfaces, or cylindrical structures [23]. During the image registration, these extracted features in one image are matched to the ones in another image, either by their appearance similarity or by geometric closeness. During the matching, correspondences are formed between features in the two images. The transformation is estimated from an objective function based on a geometric distance measure. The reader can be referred to references [24–28] for more details.

Let I_p and I_q be two images. The set of features in I_p is \mathcal{G}_p with features $\mathbf{p} \in \mathcal{G}_p$, whereas features \mathbf{q} are in the feature set \mathcal{G}_q in the image I_q . Note that \mathbf{p} represents both the feature and its location. The correspondences (\mathbf{p}, \mathbf{q}) form a set \mathcal{C} . The transformation function from I_p to I_q is $\mathbf{T}(\mathbf{p}, \theta)$, with parameter vector θ . Following the parameterized transformation functions, the geometric distance measure between two features \mathbf{p} and \mathbf{q} is denoted as $d(\mathbf{p}, \mathbf{q})$. Given a correspondence set \mathcal{C} , the least-squares objective function for feature-based methods can be written as

$$\min_{\theta} \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{C}} d^2(\mathbf{q}, \mathbf{T}(\mathbf{p}; \theta)). \quad (2.1)$$

In most cases, however, because the correspondence set \mathcal{C} is either unknown or contaminated by a large fraction of outliers, the estimation of (2.1) cannot be carried out by conventional optimization methods. Instead, alternative ways of matching and estimation have been developed to handle this issue in an iterative fashion by finding the closest points and estimate the transformation in each iteration. The next subsections discuss that method in detail.

2.2.1 Iterative Closest Point Method

The first class of feature-based methods contains the Iterative Closest Point (ICP) algorithms [29] and its variations. Starting from an initial estimate, ICP iteratively (a) maps points (features) from one image, I_p , to the other image, I_q , (b) finds the closest point on image I_q for each mapped point, and (c) re-estimates the transformation using these temporary correspondences in the objective function in (2.1).

In other words, the correspondence set \mathcal{C} is obtained from forming correspondences with the closest points. The convergence proof of ICP has been shown by Besl and McKay [29] using a Euclidean distance metric for d in (2.1). In the registration of range data, sensor data points (together with estimated normals) are the features commonly used. In the alignment of intensity images, feature extraction techniques are applied to obtain features to be matched in ICP [24].

Extraction of Harris corners is commonly seen in the literature [30, 31]. Can et al. [24] extracted vessel centerline points (edge-like features) for the registration of retinal images. There are two commonly-known problems with ICP: (1) ICP has a narrow domain of convergence and therefore must be initialized relatively accurate, and (2) feature extraction can be unreliable and overly sensitive to the choice of parameters and the image content.

The distance measures used in (2.1) are usually associated with the types of features being extracted. For corners or points, a natural distance measure is Euclidean distance [32, 33]. However, the situation is different for points sampled on a curve on a 2D image. As they are discrete samples from a continuous curve, we are not concerned with how far a point moves along the curve, but how far a point is off the curve. Hence, we want to compute point-to-curve distances.

One efficient way to achieve this is to *linearize* the curve in a small neighborhood around the sample point. The point-to-curve distance is then measured by the distance along the normal direction of the straight line [24]. The same linearization is also applied to points sampled on a surface in 3D image [34]. Such a linearization can be viewed as the first order approximation to the true distance function. Pottmann et al. [35] proposed a second-order approximation that takes local curvature into account by fitting a local tangent to the curve.

2.2.2 EM-ICP Method

Several papers (*see e.g.*, [36, 37]) have proposed Expectation Maximization (EM) [38] algorithms for image registration [39, Ch. 3], where multiple correspondences per feature are simultaneously considered. In this approach, correspondences are considered as soft, each associated with a probability of the correspondence being correct. The EM algorithms are carried out by alternating between the computation of this correspondence probability (expectation) and the estimation of transformation parameters (maximization). In [40], multiple correspondences are modeled with a Gaussian mixture model with an isotropic covariance. Later on, Chui and Rangarajan [41] exploited the EM algorithm along with simulated annealing and a doubly-stochastic normalization over the correspondence matrix to solve for free deformations [37]. This approach improves the robustness and reliability of ICP.

2.2.3 Keypoint Matching With Invariant Descriptors

In contrast to ICP where matching is based on geometric distance, keypoint-based methods apply matching in appearance similarity space. The term *keypoint* refers to a combination of an interest point (or region) and an invariant descriptor which describes the intensity characteristics in a small neighborhood around the interest point. The latter part, the descriptor, is the key difference between the definitions of keypoint and feature. Specifically, a descriptor is designed to be invariant to some lower-order geometric transformation, *e.g.*, similarity or affine, and usually additionally invariant to linear intensity changes.

Keypoint methods have received growing attention because of their demonstrated ability to tolerate low image overlap and image scale changes [32–34, 42–44]. These methods start with keypoint detection and localization followed by computation of the descriptor that summarizes the image around the keypoint. For instance, in [45], a multi-stage processing scheme is presented using a point-based registration algorithm to initialize the unknown motion field, followed by a robust M-estimator to refine the registration accuracy. Existing keypoint extraction algorithms are based on approaches ranging from the Laplacian of Gaussian operator [42], information theory [46], Harris corners [43], and intensity region stability measures [47]. A detailed comparison can be found in [48]. Region descriptors proposed are based on steerable filters [49], moment invariants [50], shape context [51], image gradients [8] and Haar wavelets [52]. A comparison of descriptors can be found in [53].

As the descriptors are invariant to geometric transformation [54], they can be compared directly by computing a similarity score between them. The keypoint matching between two images is carried out by computing similarity scores for all possible keypoint pairs and selecting the most similar ones to form the initial

set of putative matches. This brute-force matching requires $O(N^2)$ time complexity for N keypoints. Additional data structures can be used to speed up this matching process. In [42], a $k-d$ tree is used to store keypoints in one image and the matching is carried out by the nearest neighbor lookup in the $k-d$ tree. By doing so, the time complexity of matching drops to $O(N \log(N))$. Subsequently, this set of putative matches is used to estimate the transformation. However, the matching process is not well-conditioned in the sense that there are inevitably some keypoints that cannot be matched, and for some others multiple matches with the same similarity are found. This yields many spurious matches in the initial set (it is not uncommon that fewer than 25% are correct). Common estimation techniques, such as least squares, least median, or M-estimators [45] perform badly with this large fraction of outliers [8]. Instead, local descriptors, such as SIFT, Harris, *etc.* are often used to eliminate mismatches and to robustly estimate the transformation (see *e.g.*, [55, 56]). As well, RANDOM SAMPLE CONSENSUS (RANSAC)-based methods typically present robustness against outliers (see *e.g.*, [57–61]).

It is worth noting that the optimization is carried out by minimizing the geometric distance between transformed points, as shown in (2.1). Therefore, the methods based on keypoint extraction and matching are still considered as feature-based methods. One may argue that the matching between keypoints is conducted using the invariant descriptor, which is a similarity measure based on local intensities. However, because the objective function does not depend on intensity differences, these methods do not belong to the intensity-based category. Methods based on keypoint extraction followed by similarity matching in

the feature space have been applied to 2D image registration and mosaic construction [42], fundamental matrix estimation [47], object recognition [8], location recognition [62], 3D registration [63] and simultaneous localization and mapping [64, 65].

Although these approaches have robustness to outliers, they can be highly sensitive to the feature extraction process, especially in the presence of global and local variant illumination changes and/or partially missing data due to occlusions [9]. Another major problem of feature-based approaches is that the sparseness of the features tend to make higher local geometric registration error, especially if the features are less dense in a specific part of the input images [52]. Thus, feature-based methods are not considered in this thesis.

2.3 Intensity-based Methods

Intensity-based methods are those approaches that try to find the best transformation parameters which describe the alignment between two input images to achieve pixel-to-pixel matching generally. These methods avoid some of the pitfalls of feature-based methods, such that the former make optimal use of the information available in image alignment, since they measure the contribution of every pixel in the image. Intensity-based methods also include two main categories: Fourier-based and correlation-based [9].

2.3.1 Fourier-based Registration Methods

Lucchese and Cortelazzo [66] present a rotation estimation method that is based on the property that the magnitude of the Fourier transform of an image and the mirrored version of the magnitude of the Fourier transform of a rotated image

have a pair of orthogonal zero-crossing lines. The angle that these lines make with the axes is equal to half the rotation angle between the two images. The horizontal and vertical shifts are estimated afterwards using a standard phase correlation method.

Foroosh et al. [67] show that the signal power in the phase correlation corresponds to a polyphase transform of a filtered unit impulse. They applied this idea to obtain the sub-pixel part of the shifts. The shift and rotation parameters can be estimated separately, because a shift only affects the phase information, while rotation affects both phase and amplitude of the Fourier transform.

Irani et al. [68] present a multi-resolution approach to compute multiple occluding motions using an iterative method. Other methods compute an optical flow field as in [69] and obtain therefore a motion vector for every pixel of the image [70]. Methods based on the correlation between shifted images can be performed either in spatial or in frequency domain. Such methods are based on the fact that the correlation between two images is maximum if they are correctly aligned. In the frequency domain, a correlation can be computed efficiently using a multiplication of the two spectra. A spatial correlation-based method is presented by Zokai and Wolberg [71].

Vandewalle et al. [72] use the property that a shift in the space domain is translated into a linear shift in the phase of the image's Fourier Transform. Similarly, a rotation in the space domain is visible in the amplitude of the Fourier Transform. Hence, their motion estimation algorithm computes the images' Fourier Transforms and determines the 1-D shifts in both their amplitudes and phases. One advantage of this method is that it discards high-frequency components, where aliasing may have occurred, in order to be more robust.

A planar shift can be described in the frequency domain as a linear phase shift,

while a rotation in the spatial domain corresponds to a rotation over the same angle in the frequency domain. On the other hand, the horizontal and vertical components of the shift can be separated by a decomposition value [73]. This frequency domain approach has the advantage that continuous values can be estimated for the shifts. However, it suffers from phase wrapping.

Pham [1] proposed both a gradient-based shift estimator and 2D projective estimator. However, through an iterative process, the biases of both previous estimators can be removed. Hence, the iterative estimator is not only unbiased, but is also optimal since their variances reach the Cramer-Rao bound.

Although the frequency-based registration methods are limited to 2-D rigid transformation, they can be efficiently exploited if the input images are not under-sampled, and their motion can be approximated and computed as linear phase differences between the two images as in [72, 74–76].

2.3.2 Correlation-based Registration Methods

Lucas and Kanade [77] proposed an optical flow registration algorithm, which has become one of the most commonly cited computer vision papers on the subject. Twenty years later, Baker and Matthews [78] presented an overview of the Lucas-Kanade algorithm, its extensions, and experimental comparisons. These methods aim at finding a transformation function such that after the application of the transformation function, the intensity differences between corresponding pixels are minimized. These methods often work in an iterative manner: warp one image on to the coordinate system of the other using the current transformation estimate, compute an update of the transformation from the image gradients or optical flow vectors, and repeat these steps until the transformation estimate converges. A simple and canonical measure used in intensity-based

methods is the sum of squared difference (SSD) between the warped image and the fixed image [2]. The objective function using the SSD measure is defined as

$$\min_{\theta} \sum_{\mathbf{x}} [I_q(\mathbf{T}(\mathbf{x}; \theta)) - I_p(x)]^2, \quad (2.2)$$

where $\mathbf{T}(\mathbf{x}; \theta)$ is the transformation function that maps coordinates of I_p onto I_q . This objective function is non-linear because it involves the image functions I_p and I_q that map coordinates to intensities. To solve this non-linear problem, the minimization is often carried out by gradient descent methods in the following way. Assuming the current estimate, (2.2) can be linearized by performing a first order Taylor expansion on $I_q(\mathbf{T}(\mathbf{x}; \theta))$ around $\hat{\theta}$ and applying the chain rule on derivatives, we obtain

$$\sum_{\mathbf{x}} \left[I_q(\mathbf{T}(\mathbf{x}; \hat{\theta})) + \nabla I_q \frac{\partial \mathbf{T}}{\partial \theta} \Delta \theta - I_p(\mathbf{x}) \right]^2. \quad (2.3)$$

Note that, in this Taylor series expansion, only the update $\Delta \theta$ is unknown and everything else is known. Thus, it becomes a normal linear least-squares problem and the update solved accordingly [25]. These methods require good initialization since the capture range is very small. As well, the biggest disadvantage of these methods is that they have a limited range of convergence. Therefore, they can be used in a coarse-to-fine estimation framework [21] (sometimes it is named a multi-resolution framework). It has been shown that these methods are more effective and have a broader capture range with multi-resolution settings. In addition to multi-resolution, an initialization method using a coarse-to-fine search on transformations has been shown adequate in [10, 11].

By definition, SSD is sensitive to intensity variations, such as shutter speed

change or illumination changes. As can be seen from the objective function defined in (2.2), illumination changes in one image would alter the shape of the objective function and affect the minimization accordingly. This limitation is a result of a lack in the existing image registration models for explicit photometric and/or illumination correction parameters. That point will be discussed in detail in Chapter 3. In the section that follows, motion field representations which provide structural constraints are discussed.

2.4 Motion Field Representation

This section discusses the parametric and nonparametric motion models.

2.4.1 Parametric and Nonparametric Motion Models

Motion field representations [79] may be divided into two categories; parametric and nonparametric models. Each has distinct advantages and disadvantages as follows.

- Nonparametric motion field models

In nonparametric motion field models, a representation of the motion field composes of a finite set of points in the 2D image plane indexed by \mathbf{x} . It is common to choose the set of points to correspond with the discrete image sampling grid which is uniformly-spaced. The advantage of this approach is that motion fields may be represented arbitrarily and may be also interpolated to yield values between sampling points.

The main disadvantage of the nonparametric representation is that it requires the estimation of a large number of motion parameters. However,

without additional assumptions, this may be impossible due to a lack of available observation constraints.

- Parametric motion field models

On the other hand, an alternative approach is to use a parametric motion model which represents the motion field over some region of the image plane. Parametric models are typically continuous functions of the spatial location \mathbf{x} . Common parametric motion models use from 2 to 12 parameters. Once the parameters of the model are determined, the model may be evaluated at any location \mathbf{x} within the region of support, thus there is no need for interpolation. Parametric models have the advantage of requiring relatively few parameters to describe a large region of the motion field. Data from larger regions of the image may be aggregated when estimating the model parameters. Since the number of model parameters is small, this tends to yield more reliable estimates.

Parametric models have drawbacks as well. It is not possible to represent arbitrary motion fields using commonly used parametric models without increasing the number of model parameters, thus being comparable with nonparametric models. Finally, for general motion fields, estimation of the region of support of the parametric model can be very difficult.

In summary, nonparametric models represent the motion field by samples. Parametric models represent the motion field by parameterized regions as mentioned in the next section.

Table 2.1: The hierarchy of 2D parametric motion models [1].

2D Model	Degrees of freedom	Coordinate transformation	Meaning
Translation	2	$x' = x + t_x$ $y' = y + t_y$	2D translation
Euclidean	3	$x' = x \cos \theta - y \sin \theta + t_x$ $y' = x \sin \theta + y \cos \theta + t_y$	Translation + rotation
Similarity	4	$x' = sx \cos \theta - sy \sin \theta + t_x$ $y' = sx \sin \theta + sy \cos \theta + t_y$	Euclidean + isotropic scaling
Affine	6	$x' = a_{11}x + a_{12}y + t_x$ $y' = a_{21}x + a_{22}y + t_y$	Similarity + anisotropic scaling + skew
Perspective	8	$x' = (m_1x + m_2y + m_3)/D$ $y' = (m_4x + m_5y + m_6)/D$ $D = m_7x + m_8y + 1$	Affine + keystone + chirping
Optic flow	∞	$x' = x + v_x(x, y)$ $y' = y + v_y(x, y)$	Free-form motion

2.4.2 Common Motion Models

In many cases, the motion between different views of the same scene can be described by one of the parametric models presented in Table 2.1.

2.4.3 Regions of Support for Motion Models

Those parametric models, manifested in Table 2.1, apply to various partitions [79] (sets of regions) of the image plane. Denoting the image plane by \mathcal{R} , a *partition* is a set of regions $\{\mathcal{R}_i\}_1^N$ such that $\bigcup_{i=1}^N \mathcal{R}_i = \mathcal{R}$ and $\mathcal{R}_i \cap \mathcal{R}_j = \phi, \forall i \neq j$.

- Global models

For global models the partition consists of a single region $\mathcal{R}_1 = \mathcal{R}$. That region of support for global motion models is the entire image plane. Global models are particularly useful for computer vision, image stabilization, camera calibration, video compression, and satellite imaging.

- Block-based models

In block-based models the partition regions \mathcal{R}_i are equal-sized rectangular blocks. A parametric motion model applies for each block. Block motion models can be applied utilizing the affine or projective model to each block. However, block-based models are poorly suited to describe general motion fields adequately as the motion model is fixed for all locations within each block, thus inadequately representing motion field discontinuities which may be located within the block itself.

- Generalized block models (meshes)

A generalization of the fixed size, regular block partition uses adaptive block-based meshes in which the regions \mathcal{R}_i are triangles or blocks of various sizes. The meshes are selected to model the intensity and/or motion structures within the image sequence. Generalized block models have the advantage of a relatively small number of regions as compared with fixed block methods.

- Region-based models

These regions $\{\mathcal{R}_i\}_1^N$ may take on arbitrary shapes in contrast to regularly shaped regions. This leads naturally to the question of how the regions are chosen. Since each region $\mathcal{R}_i \subseteq \mathcal{R}$ represents the region of support of a motion model, \mathcal{R}_i is defined to cover those parts of \mathcal{R} where the model is appropriate. Thus, the partition is a *segmentation* of the motion field into regions. Unfortunately, determining the regions \mathcal{R}_i given the image sequence is a difficult process. Region-based motion representations often provide efficient motion representations [80].

- Points

The nonparametric models can be considered as a special case of parametric models if the region of support for each motion model is to be a single point. Assuming that the image plane \mathcal{R} is sampled on a finite lattice Λ consisting of N unique sampling locations, the partition of Λ consists of the set of N locations $\{\mathcal{R}_i\}_1^N$. A parametric motion model is then associated with each location \mathcal{R}_i . Thus the total number of parameters describing the motion field on Λ is the same as the number of sample points.

2.5 Relating Motion and Image Intensities

In this section, the relationship between the apparent motion and the observed image intensities is discussed. One of the most common assumptions made when attempting to determine optical flow is that the image intensity along a motion trajectory remains constant [11]. This is referred to as the *brightness constancy assumption* (e.g., [71, 81–83]), which is utilized in several motion estimation techniques.

Although a number of researchers [10–13] have incorporated global illumination correction into the geometric image registration problem, these approaches to modeling intensity variations are valuable for slowly changing intensity differences. In other words, these approaches don't easily allow arbitrarily-shaped illumination variations which is true for many imaging scenarios (see e.g., Fig. 2.1(a,b); canyon pair captured at widely varying times of day, and Fig. 2.1(c,d); LANDSAT satellite pair taken at two different times of year).

In principle, we should differentiate between photometric corruption and illumination variations from an image registration perspective. Photometric corruption takes place due to issues that relate to the camera itself, such as exposure time, camera response function, sensor noise, *etc.* Illumination variations are concerned with the scene itself, such as capturing two images for the same scene with perfect camera at two different times in a day assuming no object motions in the scene (*i.e.*, static scene), resulting in different shadows, *etc.* Although in this dissertation we focus on the impacts of illumination variations only, a brief background on the photometric correction-based image registration approaches is presented.

Many approaches have been proposed for geometric image registration in the presence of photometric corruption. There are three possible approaches to solve this problem. First, images are photometrically registered using an algorithm that is insensitive to geometric misalignments, followed by geometric registration. Grossberg and Nayar [14] propose an intensity mapping function algorithm that is based on histogram specification and provides intensity mapping for an image pair when there is no saturation or significant geometric misalignment.

Second, images are geometrically registered using an algorithm that is exposure-insensitive, followed by photometric registration. In [15], such a photometric changes-insensitive motion estimation algorithm based on the Lucas-Kanade technique is proposed to estimate motion vectors at each pixel. Although this algorithm can be used to estimate a large and dense motion field, it requires knowing the camera response function in advance.

Third, geometric registration and photometric and/or illumination variation

correction parameters are jointly estimated. In essence, these registration approaches have a limitation of being operated either on a global illumination model (GIM) or on an affine illumination model (AIM) that do not easily allow arbitrarily-shaped regions. For instance, Aguiar [11] proposed GIM to relate the intensity levels of an image pair. Altunbasak et al. [13] presented AIM with triangular or quadrilateral region support. Periaswamy and Farid [10] proposed a global illumination model within pre-defined neighborhoods with an imposed smoothness constraint between neighborhoods. In [12], a dual inverse compositional algorithm is used based on an assumption that the geometric and photometric transformations can be replaced, thus impeding the use of explicit local photometric transformations. The author of [12] models the photometric changes in a gain and offset fashion, when dealing with gray images, quite the same as in GIM.

One can see that the approaches in [12, 13] tend to limit the region with illumination changes to a predefined shape, which does not correspond to real imaging scenarios where this region can take an arbitrary shape. Hence, we will consider the global illumination model in [10, 11] because it assumes that the image is one region with its own gain and offset scalars. This assumption can be expanded to segment an image into many regions with arbitrarily-shaped illumination variations so as to match the real imaging scenarios.

The resulting registration from these models tends to fail when dramatic and disjoint intensity shifts over arbitrarily-shaped regions exist. Another limitation is that the occluded areas, which are not considered in their models, negatively affect the geometric registration accuracy. Therefore, we will focus on the intensity-based geometric registration jointly with illumination correction of images with arbitrarily-shaped local illumination changes, due to large shadow differences,

such as from time of day capture. These illumination changes are extracted through an iterative framework to allow the geometric registration to converge to sub-pixel accuracy.

2.6 Applications

This section is an overview on some of the applications that require high sub-pixel registration accuracy.

- **Computer Vision:**

Computer vision applications include automatic change detection for security monitoring, motion tracking, target template matching with real-time images, automatic quality inspection, and image mosaicking [2].

- **Remote Sensing:**

Many applications can be involved under remote sensing, such as panchromatic images, better spatial resolution, multispectral images with better resolution, radar images independent of cloud cover and solar illumination, monitoring of global land usage, landscape planning, mosaicking of images of the surveyed area and registration of aerial or satellite data [2].

- **Change Detection:**

Typical change detection consists of operations that compare information pixel-wise in layered data that has a common spatial reference frame. Misalignment of the layers results in some errors from mismatched locations. Several authors have studied the implications of misalignment for change

detection. Dai and Khorram [84] explains how errors of over and underestimation of change detection occur and how correct responses are diminished. A statistical framework for computing change detection error under given data is derived by Roy [85]. Corrective processing of detected change is presented by Bruzzone et al. [86], [87] with adaptive decision making using change vector analysis, and by Beauchemin and Fung [7] using adaptive median filtering to post-process detected change.

- **Medical Imaging:**

Medical image registration has emerged in many clinical applications including diagnosis, longitudinal studies, surgical planning, and the need for registration across different imaging modalities. Some examples of medical imaging applications are monitoring of healing therapy, monitoring of tumor evolution, combination of sensors recording the anatomical body structure like magnetic resonance image (MRI), ultrasound or computer tomography (CT) with sensors monitoring functional and monitoring metabolic body activities with positron emission tomography (PET), single photon emission computed tomography (SPECT) or magnetic resonance spectroscopy (MRS). Medical image registration, however, still has many challenges such as the nonlinear nature of the images used, and the difference in resolution of images acquired from different modalities (*e.g.*, MRI, CT, PET) [10].

- **Super-resolution:**

The super-resolution (SR) problem is defined as the process of combining multiple frames of aliased low-resolution (LR) images of the same scene to

form a single high-resolution (HR) image [27]. Such a combination of information from multiple frames is not trivial. There are two main issues that need to be solved in such an ill-posed SR problem. First, all the input images need to be correctly aligned with each other on a common grid; that is called *image registration*. Next, an accurate and sharp image has to be formed from the gathered information; that is called *image reconstruction or fusion*. If one of these two steps is inaccurate, the resulting HR image will be degraded, and no gain in actual resolution may result. Readers can be referred to some of the presented SR techniques in [72, 74, 75, 88–98].

2.7 Summary

In this chapter, we have shown that the image registration process plays a central role in many application. Also, we mention that feature-based registration approaches have robustness to outliers as well as some robustness against intensity variations. However, these approaches are highly sensitive to the feature extraction process. Also, those approaches suffer from the sparseness of features, thereby impacting geometric registration accuracy.

Contrarily, we have exhibited the capability of intensity-based registration approaches of being motion model independent. However, illumination variations resulting due to different times of capture is limitedly considered in the literature for *global* or *predefined-shapes*. In the next chapter, we propose a new intensity-based registration model that can handle arbitrarily-shaped local illumination variations.

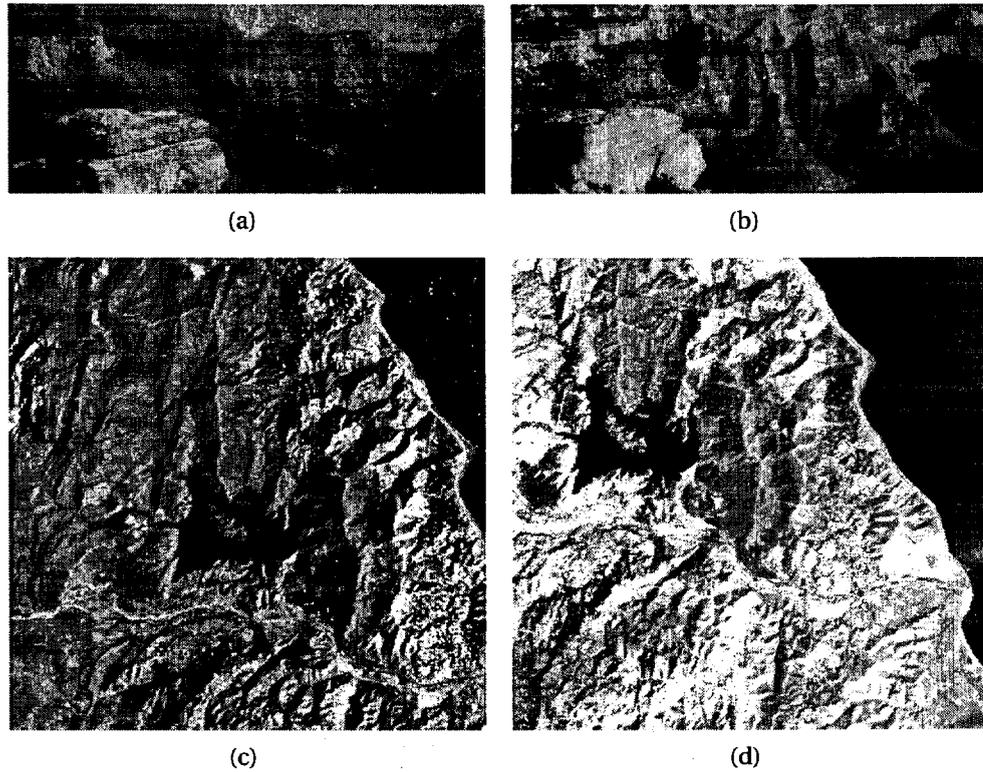


Figure 2.1: (a,b) 120×300 natural canyon pair taken at different times of day, (c,d) 400×400 real LANDSAT pair captured at different times of year.

Chapter 3

Proposed Registration Model Under Local Illumination Variations

3.1 Introduction

In the previous chapter, we have shown that competing registration models assume either global or affine illumination correction, and do not consider arbitrarily-shaped local illumination variations. The challenge is how to cope with images having such illumination changes of arbitrary shapes. One solution is to assign gain and offset to each illumination region. We propose this remedy, within an intensity-based registration model, to improve the GRP jointly with illumination correction, thereby enhancing the performance of any subsequent processing.

This chapter is organized as follows. In Section 3.2, an image formation model is presented for that assumed in the thesis. In Section 3.3, we propose a registration model that deals with images having arbitrarily-shaped local illumination variations, ASLIV, to obtain illumination correction jointly with improving

the sub-pixel geometric registration precision. In Section 3.4, we present an approach using the proposed ASLIV model with the least-squares estimator in an iterative coarse-to-fine scheme, referred to as the LS-ASLIV approach. In Section 3.5, the image data sets, used in the experiments, are fully described. In Section 3.6, we state the measures that evaluate the performance of the proposed approaches with competing approaches. In Section 3.7, an analysis on the stopping conditions is provided. The presented approach is summarized in Section 3.8. In Section 3.9, the experiments are developed and compared to competing approaches. Finally, an analysis of the LS-ASLIV approach is given in Section 3.10.

3.2 Image Formation Model

In this section, we give an overview on the image formation model that is assumed in the thesis. To present the image model, we should describe the circumstances related to capturing images for a scene, such as extrinsic and intrinsic camera issues, lighting issues, *etc.* as follows:

1. Extrinsic Camera Issues:

- The camera used is assumed to be moving.
- The camera is far from the scene.
- The distance of the camera from the scene is very large compared to the relief depth of the scene (*i.e.*, an affine motion model would be reasonable to represent the motions between the captured images with negligible parallax).
- The camera captures some images for the scene under different illumination conditions.

- A narrow base-line is also assumed.

2. Intrinsic Camera Issues:

- The camera follows the pinhole model.
- The camera is assumed to be perfect, so that photometric corruptions and lense distortions are assumed to be corrected.

3. Lighting Issues:

- Far-field light sources are assumed w.r.t. the scene.
- The scene exhibits a Lambertian model of illumination on its surfaces.

4. Scene Issues:

- The scene potentially has objects with 3D structure.
- The scene is static from the perspective that there are no moving objects.
- The scene is dynamic from the perspective of changes in illumination.

An example of a real image pair, following the constraints above, is shown in Fig. 2.1(a,b). This figure shows two images of the same canyon taken at two different times of day from different camera positions with the same exposure.

3.3 The Proposed Model

In this section, we propose an image registration model to achieve both improving the GRP and illumination correction following the image formation model shown in Section 3.2.

Consider two $N \times M$ input images, I_1 and I_2 , following the image formation model shown in Section 3.2. Under GIM [10, 11], the registration model that relates the intensity levels of I_1 and I_2 was proposed as

$$I_2(\mathbf{x}_2) \simeq bI_1(\mathbf{x}_1) + c, \quad (3.1)$$

where b and c are gain and offset scalars, respectively, and the $\mathbf{x} = (x, y)$ are a spatial coordinates for each image. While GIM allows the modeling of global illumination changes, GIM does not allow for local illumination variations, such as seen in with the shadow differences in Fig. 2.1(a,b).

Altunbasak et al. [13] introduced an extension to GIM by introducing affine transformations for the gain and offset as

$$I_2(\mathbf{x}_2) \simeq (b_1x + b_2y + b_3)I_1(\mathbf{x}_1) + (c_1x + c_2y + c_3), \quad (3.2)$$

where b_1 , b_2 , and b_3 as well as c_1 , c_2 , and c_3 are scalars. We refer to this model as AIM. You may notice that each pixel has both gain and offset values that are dependent on its position, though, some level of local illumination is introduced by tessellating the illumination correction into triangular or quadrilateral regions for independent analysis [13]. For an image such as the canyon in Fig. 2.1(a,b), coarse tessellation would leave numerous small uncorrected illumination regions that would still degrade the GIR, while conversely shrinking the tessellations would result in a large number of subregions making convergence less tractable.

Following the strategies for GIM [10, 11] and AIM [13], we propose to instead

generalize the registration model as

$$I_2(\mathbf{x}_2) \simeq B(\mathbf{x}_1)I_1(\mathbf{x}_1) + C(\mathbf{x}_1) \quad \forall \mathbf{x}_i \in \text{FOV}_1 \cap \text{FOV}_2, \quad (3.3)$$

where $\mathbf{x}_i = (x_i, y_i)$ is a pixel location and $\{I_i, B, C\} \in \mathbb{R}^{N \times M}$ for $i \in \{1, 2\}$, and I_1 and I_2 are the sensed and reference images, respectively. The illumination variations, B and C , in (3.3) are modeled as matrices such that

$$B = \begin{bmatrix} \beta_{11} & \dots & \beta_{1M} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \dots & \beta_{NM} \end{bmatrix}, \quad C = \begin{bmatrix} \zeta_{11} & \dots & \zeta_{1M} \\ \vdots & \ddots & \vdots \\ \zeta_{N1} & \dots & \zeta_{NM} \end{bmatrix}, \quad (3.4)$$

where β_{nm} and ζ_{nm} are gain and offset scalars, respectively, $1 \leq n, m \leq N, M$. The multiplication in (3.3) is element wise over \mathbf{x}_1 . Also, to ensure that the relation in (3.3) holds only for the overlapping fields of view (FOV) for I_1 and I_2 , we also explicitly specify that (3.3) is only valid for \mathbf{x}_1 and \mathbf{x}_2 falling in the intersection of the two fields of view.

As well, the motion in (3.3) could be modeled as

$$\mathbf{x}_2 = \Upsilon(a_1, \dots, a_n; \mathbf{x}_1), \quad (3.5)$$

where $\Upsilon(\cdot)$ refers to a general motion model. The corresponding motion parameter vector field can be denoted by $A = [a_1, a_2, \dots, a_n]^\top$. The primary difference between the proposed model and those presented in [10, 11] and [13]¹ is that B and C are matrices as opposed to scalars. With B and C as matrices, arbitrary illumination changes are possible at a pixel-by-pixel level between I_1 and I_2 .

¹Since we focus only on illumination changes, the proposed model excludes some other components considered in [13], such as lens distortion correction.

An obvious problem with B and C as matrices is that an infinite number of solutions exist for β_{nm} and ζ_{nm} relating pixels' value in I_1 to those in I_2 . Given a particular motion vector A , each pixel correspondence between I_1 and I_2 is an equation in the two unknowns β_{nm} and ζ_{nm} . To solve for any particular β_{nm} and ζ_{nm} , we would need to apply an appropriate constraint.

To constrain the solution of (3.3), we can analyze the problem as follows. Given a scene with specular reflection and diffusion due to a source light, such as the sun, some regions with reasonably distinct illumination variations are created. That case can hold for aerial and satellite imaging. By assuming these illumination changes, we can generally argue that each image I_i will contain v_i distinct levels of illumination (*i.e.*, v_i refers to the number of shading levels per input image). First, to exhibit the idea of the proposed model, we will start with the case that *a priori* known information of distinct shading levels is available. Then, a real image scenario where no *a priori* known information is available, will be discussed.

Given that *a priori* known information of distinct illumination levels is available, Fig. 3.1(a) shows an example of a masked image I_1 segmented into three distinct illumination levels: L_1^1, L_1^2 and L_1^3 (*i.e.*, $v_1=3$). Similarly, Fig. 3.1(b) depicts another masked image I_2 segmented, as well, into a different set of distinct illumination levels: L_2^1, L_2^2 and L_2^3 (*i.e.*, $v_2=3$). With a rough geometric registration for I_1 and I_2 , an absolute image difference (AID), $I_{1,2}$, is created having a set of illumination regions, R_j , such that

$$R_j = f(I_{1,2}; j), \quad (3.6)$$

where $f(\cdot)$ refers to the *a priori* known segmentation step of $I_{1,2}$ into a specific

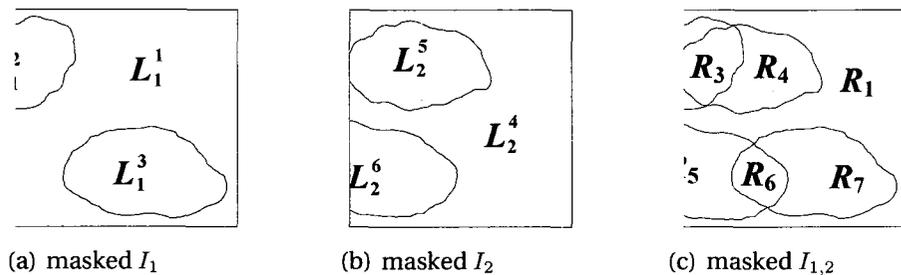


Figure 3.1: With *a priori* known information, (a,b) two different sets of distinct illumination levels of two input images (*i.e.*, $v_1=v_2=3$), yielding (c) a masked AID with seven arbitrarily-shaped regions, each has its own constant illumination (*i.e.*, $J=7$).

number of illumination regions according to J . Note that the number of resulting illumination regions, J , is *a priori* known (see Fig. 3.1-c).

It is worth noting that the sum of the number of distinct illumination levels in both images (*i.e.*, $v_1 + v_2$) does not necessarily express the number of the resulting illumination regions, J , as the latter depends on the former and the intersections among these distinct levels, L_1^h and L_2^l , $\forall h$ and l . For instance, the image pair depicted in Fig. 3.1(a) has three distinct illumination levels (*i.e.*, $v_1=3$) and so do Fig. 3.1(b) (*i.e.*, $v_2=3$), whereas the number of the resulting illumination regions, J , is seven as shown in Fig. 3.1(c). We'd like to stress that we assume each resulting illumination region, R_j , has its own gain b_j and offset c_j . Note that the proposed ASLIV model in (3.3) has piece-wise b_j and c_j .

In a real imaging scenario, the *a priori* known information on how the input images, I_i where $i \in \{1, 2\}$, are to be segmented to v_i distinct illumination levels is typically unavailable. In turn, the corresponding L_i^v , $\forall v$, cannot be obtained

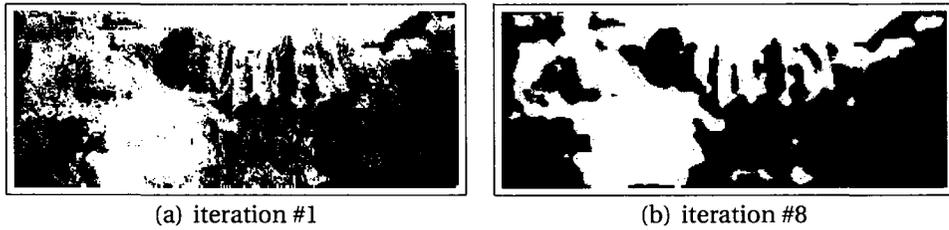


Figure 3.2: For the pair in Fig. 2.1(a,b), Binary masks (Q_1 in black, Q_2 in white) of two distinct illumination regions computed iteratively from image intensity differences at (a) 1st iteration and (b) 8st using the k -means algorithm, setting k to 2 (*i.e.*, $J=2$).

directly looking at each image on its own. We propose a simple method of finding the AID between the roughly aligned images, $I_{1,2}$, using a segmentation algorithm. For instance, the k -means algorithm [99] can be employed to segment $I_{1,2}$, setting k to the desired J (*i.e.*, J is then assumed as it is unknown). Note that $I_{1,2}$ is first determined using a rough geometric registration, then iteratively changed yielding overlapping regions R_j at each iteration, according to (3.6). For example, Fig. 3.2(a) and (b) show the segmentation of the AID of the pair shown in Fig. 2.1(a,b) at the 1st and the 8th iterations, respectively, setting k to J , where $J=2$ [18]. A descriptive scheme showing the proposed registration model using the k -means algorithm will be presented in Section 3.4.

The proposed model requires explicit segmentation at each iteration. Given the segmentation, a rough geometric registration will produce illumination regions, R_j , using (3.6) according to the information availability. For pixel domain mathematical manipulations, each R_j can be represented by a binary mask, $Q_j(x, y)$, such that

$$Q_j(\mathbf{x}) = \begin{cases} 1, & \forall \mathbf{x} \in R_j \\ 0, & \text{otherwise} \end{cases}, \quad (3.7)$$

where $1 \leq x, y \leq N, M$. Recall that each illumination region, R_j , where $1 \leq j \leq J$, has its own gain b_j and offset c_j . We can thus constrain B and C in (3.3) to these J regions as follows

$$B(\mathbf{x}) = \sum_{j=1}^J b_j Q_j(\mathbf{x}), \quad C(\mathbf{x}) = \sum_{j=1}^J c_j Q_j(\mathbf{x}). \quad (3.8)$$

In this section, we have shown how two input images, to be aligned, are segmented to J distinct illumination regions, thereby the proposed model in (3.3) is adequately constrained. In the following section, we present an approach to estimate the unknown geometric registration parameters, (a_1, \dots, a_n) , the gain parameters, (b_1, \dots, b_J) , and the offset parameters (c_1, \dots, c_J) .

3.4 Image Registration using LSE

This section exhibits an iterative approach to solve a non-linear minimization problem using the LSE in a coarse-to-fine scheme.

We need to estimate the values in the parameter vector Φ for which we use least squares estimation as

$$\hat{\Phi} = \arg \min_{\Phi} \left\{ \mathcal{L} = \sum_{\mathbf{x} \in \text{FOV}_1 \cap \text{FOV}_2} [E(\Phi)]^2 \right\}, \quad (3.9)$$

where \mathcal{L} is a quadratic cost function and $E(\cdot)$ is the residuals that can be obtained from (3.3) as

$$E(\Phi; \mathbf{x}_1) = I_2(\mathbf{x}_2) - B(\mathbf{x}_1)I_1(\mathbf{x}_1) - C(\mathbf{x}_1) \quad (3.10)$$

where the multiplication is element-wise. Note that $E(\Phi; \mathbf{x}_1)$ is an element located at position \mathbf{x}_1 in the matrix $E(\Phi)$, where $E(\cdot) \in \mathbb{R}^{N \times M}$. Assuming no object

motion in the scene, the motion vectors between I_1 and I_2 can be generally defined as

$$x_2 = x_1 + s(\mathbf{x}_1), \quad y_2 = y_1 + t(\mathbf{x}_1), \quad (3.11)$$

where $s(\mathbf{x})$ and $t(\mathbf{x})$ are the horizontal and vertical components of the motion vector, respectively. Assuming that I_1 and I_2 are continuous and differentiable functions w.r.t. their x and y , we can rewrite I_2 at the spatial location \mathbf{x}_1 in terms of its truncated 1st-order Taylor series expansion, such as

$$\begin{aligned} I_2(\mathbf{x}_2) &= I_2(x_1 + s(\mathbf{x}_1), y_1 + t(\mathbf{x}_1)), \\ &= I_2(\mathbf{x}_1) + \nabla_x I_2(\mathbf{x}_1) s(\mathbf{x}_1) + \nabla_y I_2(\mathbf{x}_1) t(\mathbf{x}_1) + h.o.t. \end{aligned} \quad (3.12)$$

Computing the higher order terms (*h.o.t.*) can be computationally expensive and can be ignored if their values are small. Thus, we rewrite (3.3) using the approximated form of I_2 to give

$$I_2(\mathbf{x}_1) + I_x(\mathbf{x}_1) s(\mathbf{x}_1) + I_y(\mathbf{x}_1) t(\mathbf{x}_1) \simeq B(\mathbf{x}_1)I_1(\mathbf{x}_1) + C(\mathbf{x}_1), \quad (3.13)$$

where $I_x(\mathbf{x}_1) = \frac{\partial I_2(\mathbf{x}_1)}{\partial x}$ and $I_y(\mathbf{x}_1) = \frac{\partial I_2(\mathbf{x}_1)}{\partial y}$. As (3.13) becomes function in \mathbf{x}_1 only, we will replace \mathbf{x} with \mathbf{x}_1 . Then, (3.10) can be written as

$$E(\Phi; \mathbf{x}) = I_2(\mathbf{x}) + I_x(\mathbf{x}) s(\mathbf{x}) + I_y(\mathbf{x}) t(\mathbf{x}) - I_1(\mathbf{x}) \sum_{i=1}^g b_i Q_i(\mathbf{x}) - \sum_{i=1}^g c_i Q_i(\mathbf{x}). \quad (3.14)$$

Assuming the 6-parameter affine motion model as presented in [10, 11, 13], which specifies the motion between I_1 and I_2 , we can then write

$$s(\mathbf{x}) = a_1 x + a_2 y + a_5, \quad t(\mathbf{x}) = a_3 x + a_4 y + a_6. \quad (3.15)$$

Then, incorporating (3.15) into (3.14), we rewrite (3.14) as

$$E(\Phi; \mathbf{x}) = I_2(\mathbf{x}) + I_x(\mathbf{x})(a_1x + a_2y + a_5) + I_y(\mathbf{x})(a_3x + a_4y + a_6) - \sum_{i=1}^g c_i Q_i(\mathbf{x}) - I_1(\mathbf{x}) \sum_{i=1}^g b_i Q_i(\mathbf{x}). \quad (3.16)$$

We can estimate Φ with the least-squares estimator in (3.9) using many algorithms, such as the Levenberg-Marquardt algorithm (LMA), Gauss-Newton algorithm (GNA) and gradient-descent algorithm (GDA) [100]. In principle, GDA has major drawbacks; i) it converges towards a local minimum in many iterations, thus it is computationally costly, and ii) computing the optimal step size per iteration can be time-consuming, and using a fixed step size often yields poor results. Both LMA and GNA have the advantage that second derivatives, which can be difficult to determine at each iteration, are not required, thus decreasing the computational time. As well, both are guaranteed to converge if the initial guess is close to the right one. However, LMA may converge slower than GNA if the initial guess of the unknown vector, Φ° , is too far from the correct answer [100]. Therefore, GNA [101] is chosen to solve (3.9) using an identity initialization that will be discussed later.

With the GNA, the $\hat{\Phi}$ is updated by Δ at each iteration g in each resolution level as

$$\hat{\Phi}_g = \hat{\Phi}_{g-1} + \Delta_g. \quad (3.17)$$

Obtaining the 1st-order Taylor series expansion of the matrix $E(\Phi)$, \mathcal{L} in (3.9) can be then rewritten as

$$\mathcal{L} \approx \sum_{\mathbf{x}} [E(\Phi_{g-1}) + \Delta_g^\top \nabla_{\Phi} E(\Phi_{g-1})]^2, \quad (3.18)$$

where $\nabla_a b = \partial b / \partial a$. Setting the gradient of \mathcal{L} w.r.t. Δ to zero, we obtain

$$-\sum_{\mathbf{x}} [E(\Phi_{g-1}) \nabla_{\Phi} E(\Phi_{g-1})] = \hat{\Delta}_g^{\top} \sum_{\mathbf{x}} [\nabla_{\Phi} E(\Phi_{g-1}) \nabla_{\Phi} E^{\top}(\Phi_{g-1})]. \quad (3.19)$$

We can rewrite (3.19) in matrix notation as

$$-\mathbf{H} \mathbf{K}^{\top} = (\mathbf{H} \mathbf{H}^{\top}) \hat{\Delta}, \quad (3.20)$$

where

$$\mathbf{H} = [H_{1,1}, H_{1,2}, \dots, H_{N,M}], \quad (3.21)$$

$$\mathbf{K} = [I_2(1,1), I_2(1,2), \dots, I_2(N,M)], \quad (3.22)$$

and

$$H_{n,m} = [n I_x, m I_x, n I_y, m I_y, I_x, I_y, -I_1 Q_1, \dots, -I_1 Q_J, -Q_1, \dots, -Q_J]^{\top}. \quad (3.23)$$

Equations (3.20) through (3.23) can be used to perform one iteration for finding a solution of $\hat{\Delta}$ to update $\hat{\Phi}$ as in (3.17). However, Q_1 to Q_J in (3.23) must also be determined. Hence, an iterative framework is proposed by first solving for the unknown vector Φ and then determining those distinct illumination regions, R_j , to ultimately obtain their corresponding binary masks, Q_j . These two steps can then be iterated to refine the parameters in $\hat{\Phi}$ and refine the arbitrarily-shaped regions and their masks. It is worth noting that segmenting the image difference between the two images to be aligned affects the geometric registration precision at each iteration. This is because some pixels of a certain region, R_p , become a

part of another region, R_q , where $p \neq q$ according to the result of the segmentation algorithm in the current iteration (*i.e.*, R_j is iteratively adapted). This segmentation issue will be addressed in Chapter 5.

The proposed registration model is referred to as ASLIV $_{n,2J}$. In the proposed model name, n refers to the number of the motion model parameters (*e.g.*, $n=6$ in case of using the affine motion model), while $2J$ refers to the illumination parameters (*i.e.*, gain and offset for each region). Accordingly, the approach presented in this chapter, which uses ASLIV model jointly with the LSE will be referred to as LS-ASLIV.

Coarse-to-fine Framework: It is worth noting that the LS-ASLIV approach employs the model in (3.3) within a coarse-to-fine framework [21] to cope with large motions and save the computational time. Also, the coarse-to-fine scheme is used to increase the geometric registration accuracy (GRA) by improving the chance of convergence. While, the focus of this dissertation is to improve the sub-pixel precision on cases where convergence indeed occurs. A Gaussian pyramid can be generally created in a non-dyadic structure. However, to simplify the implementation, we have built a Gaussian pyramid of r resolution levels in a dyadic form for input images. The registration process is initialized by Φ° at the coarsest resolution level and updated by Δ for a certain number of iterations per resolution level. The approach turns to the next finer level, once one of the following criteria occurs: either the cost function in (3.9) is updated by less than a predefined threshold, ε , or a maximum number of iterations, g , has been reached.

When the solution doesn't diverge, convergence is achieved at the last resolution level using the same criteria as above. In turn, the illumination parameters are adjusted to the resulting up-scaled Q_j in (3.7). At each iteration, the estimated Φ is used to align the sensed image, I_1 , to the reference image, I_2 , with corrected

illumination as in (3.3). In fact, since an approximate estimate is achieved by iterating at the lowest resolution, the subsequent refinement of that estimate requires only a few iterations at larger resolution levels. The geometric and illumination parameters at each level of the pyramid are accumulated yielding a single final estimate.

Fig. 3.3 illustrates a scheme of the proposed image registration approach using the k -means algorithm for segmentation. Given two input images, I_1 and I_2 , as described in Section 3.3, the proposed approach can be summarized as follows:

1. A rough registration of I_1 with I_2 is performed using GIM, AIM, or some other image registration approach, or that we have *a priori* knowledge of Φ that allows for a reasonable initial image alignment.
2. The absolute image difference, I_{21} , between the two roughly aligned images is segmented into J illumination regions. The segmentation step is obtained using the k -means algorithm [99] with which the residuals (*i.e.*, AID) are clustered into k clusters, with $k=J$.
3. The unknown vector, Φ , including the geometric and illumination parameters is estimated using (3.9).
4. With an iterative coarse-to-fine scheme, the approach stops when the stopping criteria have been matched. The resulting segments (*i.e.*, the illumination regions) are iteratively created and refined. Note that the segmentation step produces the binary mask shown in Fig. 3.2(a) at iteration #1 and another binary mask depicted in Fig. 3.2(b) at iteration #8, given that Fig. 2.1(a,b) are the input images I_1 and I_2 , respectively.

5. Otherwise, the approach turns to the next iteration or to the next finer resolution level with application of the estimated vector on I_1 .

3.5 Data Set Description

In this section, we describe the image data sets used in the experiments. The data sets include two categories: real and simulated. The first category involves a 120×300 natural canyon pair (see Fig. 2.1(a,b)), and nine real 400×400 and 600×600 LANDSAT satellite images [102] with unknown ground truth motion (see *e.g.*, Fig. 2.1(c,d)). Appendix A contains the other eight real pairs.

The second category of the data sets includes two groups of simulated image pairs. Each group has 50 pairs of particular J . The size of the images in each pair is 500×500 pixels acquired from 3000×3000 pixels IKONOS satellite images for the Pentagon and its surroundings. A “ J ” simulated data set is constructed as follows. First, we have fifty 500×500 reference images. Each *reference* image, I_2 , is geometrically deformed using *a priori* known affine transformation by a random rotation angle of up to $\pm 10^\circ$ and by a random 2-D translation of sub-pixel accuracy within ± 6 pixels, in both horizontal and vertical directions. The approach in [103] is used to perform such geometric deformations. Thus, a *sensed* image, I_1 , is produced. Then, both sensed and reference images are subjected to varying v_1 and v_2 levels of illumination. Thus, variant shaded regions are produced leading to a “ J ” simulated pair. Similarly, the rest of the fifty pairs are produced yielding a “ J ” data set. With multiple “ J ”s, the simulated data sets are used since they allow ground truth verification of the estimated parameters. In the experiments, we exploit two groups of simulated data sets defined as follows:

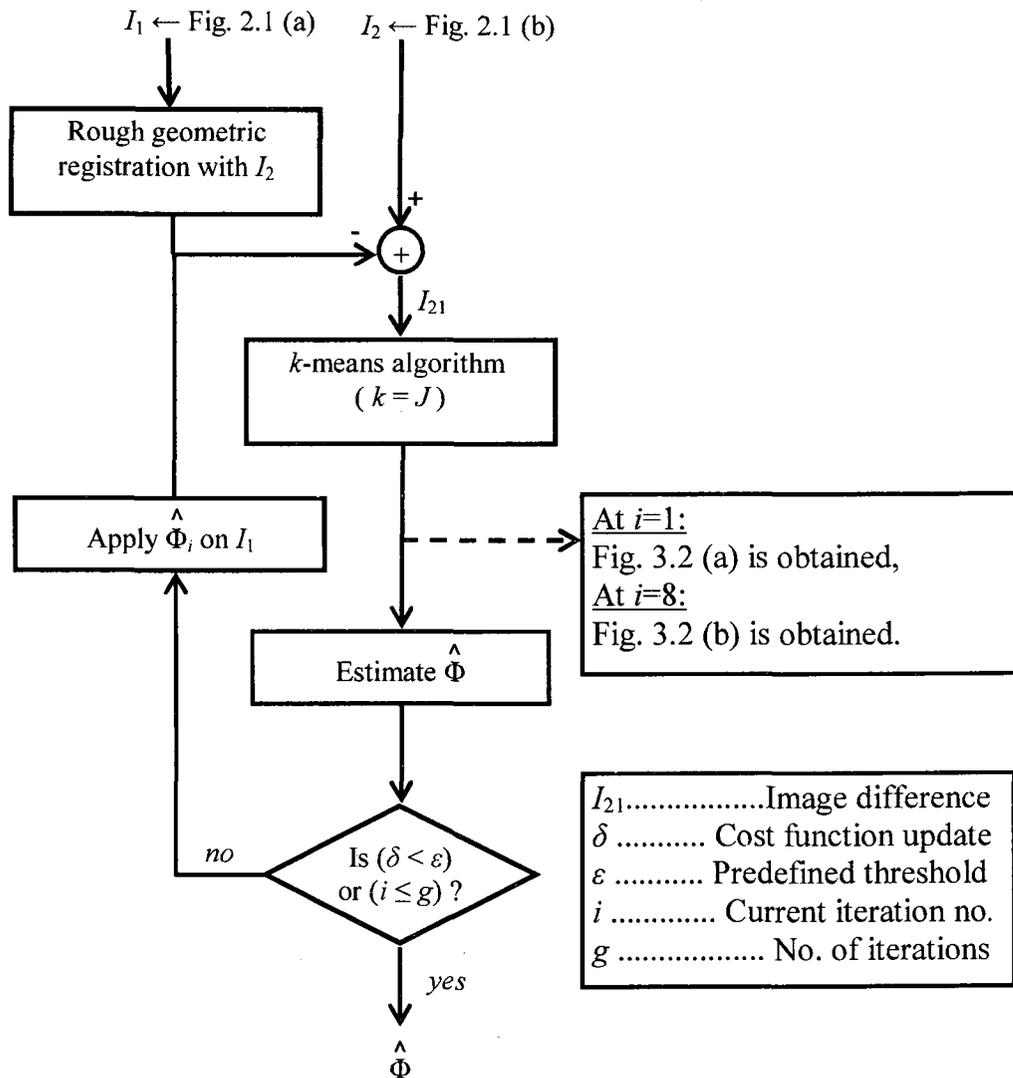


Figure 3.3: An illustrative scheme of the proposed registration approach, LS-ASLIV, using the k -means algorithm for segmentation. Given the proposed registration model in (3.3) with $k=J=2$, the natural pair depicted in Fig. 2.1(a,b) is used as an example producing the binary masks shown in Fig. 3.2(a) and (b) at iteration #1 and #8, respectively.

Definition 1 “ $J=3$ ” data set: is the fifty simulated image pairs constructed using the notations in Fig. 3.1(a,b) as follows: $L_1^2 = L_1^3 = L_2^2 = L_2^3$, and $L_1^1 \neq L_1^2 \neq L_1^3$. These settings mean that $v_1=v_2=2$, assuming no intersections between the resulting regions (e.g., Fig. 3.4(a,b)).

Definition 2 “ $J=4$ ” data set: is the fifty simulated image pairs constructed using the notations in Fig. 3.1(a,b) as follows: $L_1^2 = L_1^3 = L_2^2$, and $L_1^1 \neq L_1^2 \neq L_1^3 \neq L_2^3$. Given no intersections between the resulting regions, these settings mean that $v_1=2$ and $v_2=3$. (e.g., Fig. 3.4(c,d)).

It is necessary to realize that the LS-ASLIV_{6,2} approach (i.e., “ $J=1$ ”) corresponds to the global illumination approach in [10, 11, 13]. As well, the LS-ASLIV_{6,4} approach (i.e., “ $J=2$ ”) is a special case discussed in [18] and deals with those image pairs having only two distinct illumination regions.

3.6 Performance Evaluation

In this section, we report commonly used image quality measures to evaluate the performance of the competing approaches. For the real and simulated data sets, we determine the correlation between the overlapping area of the two $N \times M$ registered images, U and V , using

(i) structure similarity (SSIM) index [104]

$$\text{SSIM}(U, V) = \left(\frac{2\mu_u\mu_v + K_1}{\mu_u^2 + \mu_v^2 + K_1} \right) \left(\frac{2\sigma_{uv} + K_2}{\sigma_u^2 + \sigma_v^2 + K_2} \right), \quad (3.24)$$

where μ_u and μ_v are the sample means of U and V , respectively, σ_u^2 and σ_v^2 are the sample variances of U and V , respectively, and σ_{uv} is the sample cross covariance

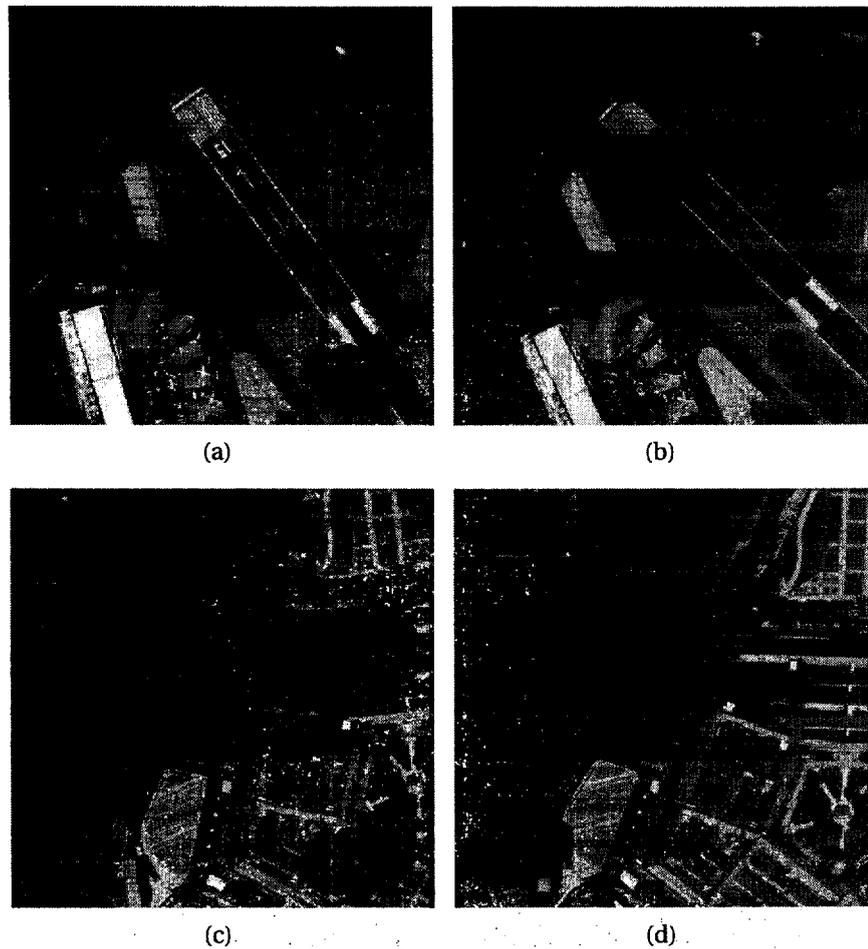


Figure 3.4: (a,b) A “ $J=3$ ” simulated pair, and (c,d) a “ $J=4$ ” simulated pair.

between U and V . The constants K_1 and K_2 are used to stabilize the SSIM index for the case where the means and variances become very small. Note that a SSIM index of 1 means a perfect match between images, while 0 means complete mismatch.

(ii) peak signal-to-noise ratio (PSNR) [105]

$$\text{PSNR}(U, V) = 10 \cdot \log_{10} \left(\frac{(\max(U))^2}{\text{MSE}(U, V)} \right), \quad (3.25)$$

where $\text{MSE}(U, V)$ refers to the mean square error between U and V , and $\max(U)$ is the maximum possible pixel value of the image U . This maximum value is set to 255, when the pixels are represented using 8 bits per sample.

(iii) normalized cross-correlation [106]

$$\text{NCC}(U, V) = \sum_{n=1}^N \sum_{m=1}^M \frac{(U_{n,m} - \mu_u)(V_{n,m} - \mu_v)}{(NM - 1)\sigma_u \sigma_v}, \quad (3.26)$$

where $U_{n,m}$ is a pixel value of image U located at position (n, m) . Note that a NCC value of 1 means a perfect match between images, while 0 means complete mismatch.

Furthermore, for only the simulated data sets, the average of absolute error (AAE) between the estimated affine parameters, $\hat{a}_1, \dots, \hat{a}_6$, and the corresponding ground truth values, a_1, \dots, a_6 , are computed for each data set. These AAEs are determined to show the mis-estimation in geometric registration using the competing models as

$$\text{AAE}(a_t) = \frac{1}{D} \sum_{d=1}^D |a_t^d - \hat{a}_t^d|, \quad t \in \{1, \dots, 6\}, \quad (3.27)$$

where D is the number of pairs per data set (=50) and a_t^d denotes the parameter, a_t , for pair # d . Note that the lower the value of AAEs, the better the performance of the approach.

3.7 Analysis on Stopping Conditions

The biggest disadvantage of the correlation-based image registration methods is that they have a limited range of convergence [9]. This section analyzes how

some critical parameters should be chosen to try to obtain the convergence with a reasonable computational time. One can notice that the proposed approach is heuristic. For instance, the proposed approach is stopped when the cost function update becomes less than a stopping threshold ($\varepsilon = 0.1$), or a maximum number of iterations ($g = 10$) is reached. Note that the proposed approach have been tested on many predefined thresholds down to 0.01 and several number of iterations up to 80 resulting in a negligible increase of 0.05%, 0.13% and 0.16% in NCC, SSIM index and PSNR, respectively, with a salient increase in the computational time by a factor of 10.

The initialization of the unknown vector, Φ , plays a central role in convergence and minimizing the computational time. With simulated data sets, the ranges of the unknown geometric and illumination parameters are already known. For each simulated pair using both “ $J=3$ ” and “ $J=4$ ” data sets, an experiment of 100 trials is performed with a random initialization using those *a priori* known ranges. This experiment yields an average success rate (ASR) in term of convergence using random initializations of 93% compared to 100% using an identity initialization, which assumes that the two input images are aligned and no illumination variations exist. As well, for the real data set, the ASR using random initializations is 8% compared to 100% using the same identity initialization. This is because the ranges of the unknown parameters are unavailable in case of real data set. Therefore, in all our experiments, we use the identity initialization as follows: Φ° is initialized either by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0]$ when using the ASLIV_{6,6}-based approaches, by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0]$ with the ASLIV_{6,8}-based approaches, or by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$ with the ASLIV_{6,10}-based approaches.

Algorithm 3.1 The proposed image registration approach LS-ASLIV

-
- 1: **Given:** two input images, I_1 and I_2 , as described in Section 3.3.
 - 2: **Required:** vector, Φ , that includes the finally estimated geometric and illumination parameters.
 - 3: **Initialize:**
 - 4: Φ ... the unknown vector by Φ^o ,
 - 5: r ... the no. of resolution levels,
 - 6: g ... the max no. iterations per resolution level,
 - 7: ε ... a predefined threshold of the cost function update,
 - 8: J ... the no. of illumination regions, and
 - 9: i ... iteration counter by 1.
 - 10: **for** $p = 1$ to r **do**
 - 11: **repeat**
 - 12: Apply $\Phi_{(i)}$ to I_1 .
 - 13: Determine the AID⁽ⁱ⁾: $I_{21} = |I_2 - I_1^{\Phi_{(i)}}|$.
 - 14: Apply the k -means algorithm to the AID⁽ⁱ⁾ to obtain the binary masks, $Q_j^{(i)}$, of the illumination regions setting k to J .
 - 15: Determine the residuals, $E(\Phi_{(i)})$, using (3.14).
 - 16: Determine the corresponding value of the cost function, $\mathcal{L}(\Phi_{(i)})$, using (3.18).
 - 17: Determine the cost function update $\delta_{(i)} = \mathcal{L}(\Phi_{(i)}) - \mathcal{L}(\Phi_{(i-1)})$.
 - 18: Determine the unknown vector's update, $\Delta_{(i)}$, using (3.20).
 - 19: Update $\Phi_{(i)}$ by $\Delta_{(i)}$ using (3.17).
 - 20: $i \leftarrow i + 1$.
 - 21: **until** ($\delta_{(i)} < \varepsilon$) or ($i > g$)
 - 22: **end for**
-

3.8 Summary on the LS-ASLIV Approach

Thus far, we have proposed an image registration model that can cope with arbitrarily-shaped regions with distinct illumination variations. The proposed registration model is cast in an iterative coarse-to-fine framework using the LSE as shown in Algorithm 3.1.

3.9 Experiments and Results

In this section, we develop the experiments to show the effectiveness of the LS-ASLIV approach as opposed to competing ones. Section 3.9.1 and Section 3.9.2 present experiments performed on simulated and real data sets, respectively, in terms of improving the geometric registration precision. We'd like to stress that we compare the LS-ASLIV approach to the global illumination approach (referred to as LS-GIM) [10, 11] and the affine illumination approach (referred to as LS-AIM) [13]. Note that the LS-GIM approach employs a global illumination model (GIM) with an LSE in an iterative coarse-to-fine framework. Also, we emphasize that the implementation of the LS-GIM approach, provided by Aguiar, can be found online at [107]. It is worth noting that the proposed approach LS-ASLIV_{6,2} corresponds to the LS-GIM approach, however, we use the code in [107].

Our implementation runs on a 2 GHz Pentium IV Core 2 Duo, with 2 GB of RAM. The LS-ASLIV approach iteratively exploits a coarse-to-fine framework with 5 resolution levels (*i.e.*, $r=5$). Consequently, the unknown vector $\hat{\Phi}$ is adequately estimated at the coarsest level for both real and simulated data sets. For the natural canyon pair, in Fig. 2.1(a,b), the resolution level is set to 3 (*i.e.*, $r=3$). The stopping threshold, ε , is set to 0.1, while the maximum number of iterations, g , is set to 10. Finally, the unknown vector, Φ , is initialized as mentioned in Section 3.7.

3.9.1 Experiments on Simulated Data Sets

The first set of experiments is performed on the “ $J=4$ ” data set using LS-GIM, LS-AIM, LS-ASLIV_{6,6}, LS-ASLIV_{6,8} and LS-ASLIV_{6,10} approaches. In fact, the main reason of performing this set of experiments is to show the performance of the

registration process when applying the proposed ASLIV model:

- i) with one less degree of freedom than an image pair has (*i.e.*, applying ASLIV_{6,6} approach on a “ $J=4$ ” image pair),
- ii) with the same degree of freedom that an image pair has (*i.e.*, applying ASLIV_{6,8} approach on a “ $J=4$ ” image pair), and
- iii) with one higher degree of freedom than an image pair has (*i.e.*, applying ASLIV_{6,10} approach on a “ $J=4$ ” image pair).

It is worth noting that the LS-ASLIV approach provides the ability to cope with non-matched data sets (*e.g.*, the LS-ASLIV_{6,8} can cope either with an “ $J=3$ ” image pair or with a “ $J=5$ ” image pair). Given that the LS-ASLIV_{6,10} approach is applied to a “ $J=4$ ” image pair. This “ $J=4$ ” image pair is originally created with J illumination regions. Each region has two separated areas with the same illumination (*i.e.*, gain and offset). The LS-ASLIV_{6,10} approach randomly chooses an illumination region, R_ν , to be segmented into two illumination regions, R_φ and R_ω $\forall 1 \leq \nu, \varphi, \omega \leq g$ (*i.e.*, each separated area of the illumination region R_ν , then, becomes a distinct illumination region yielding R_φ and R_ω). Then, the same “ $J=4$ ” image pair would behave as if it is a “ $J=5$ ” image pair.

Similarly, with the ASLIV_{6,6} being applied to a “ $J=4$ ” image pair, the ASLIV_{6,6} provides the ability to randomly choose two illumination regions, R_ι and R_ρ , to be merged into one illumination region, R_τ $\forall 1 \leq \iota, \rho, \tau \leq g$. Thus, the same “ $J=4$ ” image pair would behave as if it is a “ $J=3$ ” image pair.

Table 3.1 shows the average of absolute error of the estimated affine parameters of the “ $J=4$ ” data set, that includes 50 simulated image pairs, computed using the competing approaches compared to the ground truth values. Since the lower the value of AAEs, the better the performance of the approach, the ASLIV_{6,8} provides more precise motion estimates as opposed to the other models. We can

analyze the results of Table 3.1 as follows.

In Table 3.1-column(i), the global illumination approach (LS-GIM) does not consider that four illumination regions (*i.e.*, “ $J=4$ ”) exist in the data set used, as it assumes that the whole image is one illumination region (*i.e.*, “ $J=1$ ”). In turn, this assumption negatively affects the GRP, leading to an imprecise estimation of the geometric (affine) parameters.

In Table 3.1-column(ii), the affine illumination approach (LS-AIM) does not consider that four illumination regions (*i.e.*, “ $J=4$ ”) exist in the data set used, as it assumes that each pixel has gain and offset such that their values are dependent on their positions. In turn, this assumption impacts the GRP, leading to an imprecise estimation of the geometric parameters.

In Table 3.1-column(iii), we apply the ASLIV_{6,6} approach with one less degree of freedom (*i.e.*, “ $J=3$ ”) than an image pair has (*i.e.*, “ $J=4$ ”). This also leads to an inaccurate estimation of the affine parameters. However, the corresponding average of absolute error in this case is much smaller than those for GIM, as all illumination levels are considered but one.

In Table 3.1-column(iv), we apply the ASLIV_{6,8} approach with the same degree of freedom that an image pair has (*i.e.*, “ $J=4$ ”). All illumination regions are taken into account.

In Table 3.1-column(v), we apply the ASLIV_{6,10} approach with one higher degree of freedom (*i.e.*, “ $J=5$ ”) than an image pair has (*i.e.*, “ $J=4$ ”). This would lead to a slight imprecise estimation of the affine parameters. It is noticeable that the linear equation system in (3.20) tends to converge as b_φ and b_ω are to be closely equal and so are c_φ and c_ω as the corresponding regions, R_φ and R_ω , have the same illumination (recall that $R_\varphi \cup R_\omega \equiv R_\nu$ as mentioned above). We believe the slight mis-estimation of the affine parameters shown in Table 3.1-column(v)

Table 3.1: The AAEs ($\times 10^{-4}$) of the estimated affine parameters using the “ $J=4$ ” simulated data set, that includes 50 simulated image pairs, with the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6}, (iv) LS-ASLIV_{6,8} and (v) LS-ASLIV_{6,10} approaches with the “ $J=4$ ” simulated data set. The lower the value of AAEs, the better the performance of the approach.

Param.	(i)	(ii)	(iii)	(iv)	(v)
a_1	16.5962	13.7167	1.9854	1.8121	1.8133
a_2	13.2015	10.6395	1.8752	1.6025	1.6057
a_3	15.5246	13.2940	1.9043	1.7360	1.7364
a_4	21.0259	17.3674	2.8607	2.4261	2.4295
a_5	9472.09	7918.66	832.49	816.50	817.32
a_6	2835.12	2382.50	826.07	804.81	804.97

w.r.t. column(iv) is due to an issue of convergence settings. In this specific experiment, by decreasing the stopping threshold, ε , down to 0.01, or increasing the maximum number of iterations, g , up to 80, the resulting values of AAEs in column(v) using the ASLIV_{6,10} are much closer to those in column(iv) using the ASLIV_{6,8} with an apparent increase in computational time by a factor of 10.

Table 3.2 depicts the SSIM indexes and PSNR values, with computational time, using LS-GIM, LS-AIM, LS-ASLIV_{6,6}, LS-ASLIV_{6,8} and LS-ASLIV_{6,10} approaches for a random subset of the “ $J=4$ ” data set. These measures represent the correlation of the overlapping areas between the two registered images. It can be noticed that the measures using LS-ASLIV_{6,8} are significantly higher than those using LS-GIM, higher than those measures using LS-ASLIV_{6,6} with noticeable decrease in computational time, and slightly higher than those using LS-ASLIV_{6,10} with obvious increase in computational time as shown in Table 3.3.

For instance, Fig. 3.5(a,b,c,d) depict the normalized AIDs between the aligned images using the LS-GIM, LS-AIM, LS-ASLIV_{6,6}, LS-ASLIV_{6,8} and LS-ASLIV_{6,10} approaches, with PSNR=20.47, 27.69, 28.88 and 28.80 dB, respectively, for the “ $J=4$ ”

Table 3.2: SSIM indexes, and PSNR for the “ $J=4$ ” data set using (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6}, (iv) LS-ASLIV_{6,8} and (v) LS-ASLIV_{6,10} approaches, for a random subset of the 50 simulated pairs, with some statistics of the whole 50 pairs reported for each column. Recall that the higher the SSIM index, the better the performance of the approach.

Pair #	SSIM [104]					PSNR (dB)				
	(i)	(ii)	(iii)	(iv)	(v)	(i)	(ii)	(iii)	(iv)	(v)
1	0.8554	0.8853	0.9522	0.9671	0.9654	22.73	24.21	29.52	30.35	30.29
12	0.8292	0.8645	0.9468	0.9594	0.9593	21.37	23.15	28.24	29.19	29.11
23	0.8407	0.8742	0.9504	0.9586	0.9579	23.75	25.54	30.41	31.26	31.19
34	0.8603	0.9021	0.9528	0.9653	0.9630	23.63	25.37	29.82	30.72	30.67
45	0.8606	0.9029	0.9524	0.9699	0.9664	20.47	22.09	27.96	28.88	28.80
μ	0.8492	0.8774	0.9597	0.9641	0.9624	22.39	24.43	29.75	30.10	30.02
σ	0.0138	0.0105	0.0049	0.0032	0.0037	1.43	1.16	0.98	1.02	1.01
Min	0.8292	0.8361	0.9530	0.9586	0.9579	20.47	22.01	27.86	28.42	28.35
Max	0.8606	0.9183	0.9628	0.9699	0.9664	23.75	25.94	30.95	31.21	31.16

simulated pair shown in Fig. 3.4(c,d). The darker the AID, the lower the AID, the more accurate the geometric registration and illumination correction. As can be seen, the LS-ASLIV_{6,8} outperforms the LS-GIM, LS-AIM, LS-ASLIV_{6,6} and LS-ASLIV_{6,10} approaches.

3.9.2 Experiments on the Real Data Set

Our next set of experiments is performed on the natural canyon pair in Fig. 2.1(a,b) and the nine real LANDSAT pairs [102], *e.g.*, see Fig. 2.1(c,d). Since no ground truth verification is offered with real data sets, we should use another image quality measure rather than SSIM and PSNR. Many image quality measures can be employed, such that the correlation between the overlapping areas of the aligned images is obtained. Therefore, we use the NCC [106] in (3.26) to evaluate the performance of competing approaches in this set of experiments. We develop these experiments using LS-GIM, LS-AIM, LS-ASLIV_{6,6} and LS-ASLIV_{6,8}

Table 3.3: Computational time for a random subset of the “ $J=4$ ” data set using (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6}, (iv) LS-ASLIV_{6,8} and (v) LS-ASLIV_{6,10} approaches, with some statistics of the 50 pairs reported for each column.

Pair #	Time (sec.)				
	(i)	(ii)	(iii)	(iv)	(v)
1	33.6	37.6	42.0	52.9	63.9
12	32.9	36.2	41.3	50.7	61.9
23	30.9	33.8	42.4	52.0	65.4
34	31.6	35.3	41.2	51.9	64.1
45	30.8	33.6	43.2	55.6	66.1
μ	31.9	35.2	41.8	52.6	63.6
σ	1.22	1.06	0.85	1.86	1.75
Min	29.1	33.2	40.7	49.5	60.0
Max	33.6	37.9	43.9	53.9	67.6

approaches with the k -means algorithm for segmentation, setting k to 3 and 4, respectively, with LS-ASLIV approaches (*i.e.*, $J=3, 4$). Table 3.4 shows that LS-ASLIV_{6,8} surpasses the LS-ASLIV_{6,6}, LS-GIM and LS-AIM in terms of NCC. Recall that the higher the NCC value, the better the performance of the approach. Note that the experimental analysis mentioned in Section 3.9.1 can also hold for the results reported in Table 3.4.

For instance, Fig. 3.6(a,b,c,d) show image mosaics of the natural canyon pair in Fig. 2.1(a,b) using the LS-GIM, LS-AIM, LS-ASLIV_{6,6} and LS-ASLIV_{6,8} approaches. These mosaics are simply created using an overlaid average of the two resulting aligned images as presented in [18]. It can be shown that LS-ASLIV_{6,8} approach yields NCC=0.9384, thereby outperforming both the LS-GIM with NCC=0.8819, LS-AIM with NCC=0.8998 and the LS-ASLIV_{6,6} with NCC=0.9226. Another real example is shown in Fig. 3.7(a,b,c,d) using the competing approaches, respectively for the real image pair shown in Fig. 2.1(c,d). One can

Table 3.4: NCC for the real data set using the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6}, (iv) LS-ASLIV_{6,8} and (v) LS-ASLIV_{6,10} approaches. NCC values express the correlation of the overlapping area between the two registered images. Note that pair #A refers to Fig. 2.1(a,b), and pair #B denotes Fig. 2.1(c,d). Note that the higher the NCC value, the better the performance of the approach.

Pair	(i)	(ii)	(iii)	(iv)	(v)
A	0.8819	0.8998	0.9226	0.9232	0.9244
B	0.9642	0.9691	0.9789	0.9800	0.9815
C	0.9799	0.9813	0.9853	0.9866	0.9879
D	0.9450	0.9520	0.9644	0.9654	0.9667
E	0.9601	0.9665	0.9743	0.9752	0.9763
F	0.9719	0.9786	0.9922	0.9931	0.9938
G	0.9551	0.9594	0.9657	0.9663	0.9671
H	0.8971	0.9087	0.9229	0.9240	0.9252
I	0.9081	0.9105	0.9241	0.9281	0.9293
J	0.9542	0.9604	0.9770	0.9828	0.9840

notice that the LS-ASLIV_{6,8} approach with NCC=0.9800 outperforms LS-GIM, LS-AIM and LS-ASLIV_{6,6} approaches with NCC=0.9642, 0.9691 and 0.9789, respectively. Fig. 3.7 is gamma-corrected for better visualization ($\gamma=1.7$).

It is worth noting that Pair *F* in Table 3.4 corresponds to Test *I* presented in [45] with NCC=0.9133. One can see that the currently proposed approach with its two model versions, LS-ASLIV_{6,6} and LS-ASLIV_{6,8}, yields NCC=0.9743 and 0.9815, respectively, thereby outperforming the approach of Xu and Kasparis [45].

3.10 Analysis & Discussion

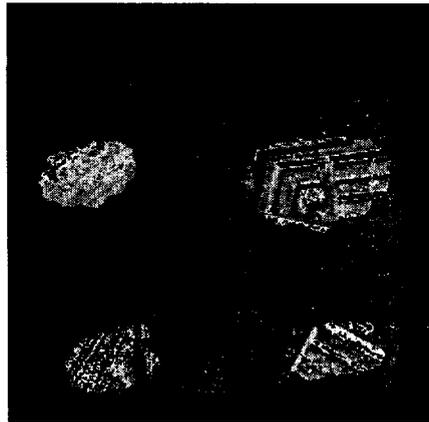
Since the main target of the research is to “*improve the sub-pixel geometric registration precision*” by minimizing the factors affecting this process, such as illumination variations, we need to analyze the LS-ASLIV approach presented in

this chapter. Unlike the LS-GIM and LS-AIM approaches, the ASLIV registration model assigns gain and offset parameters to each distinct illumination region. Then, the ASLIV model is employed with the LSE introducing the registration approach, LS-ASLIV. Although the experiments in this chapter show that the LS-ASLIV approach outperforms both the LS-GIM and LS-AIM approaches, qualitatively and quantitatively, we believe there are still some issues that negatively impact the registration.

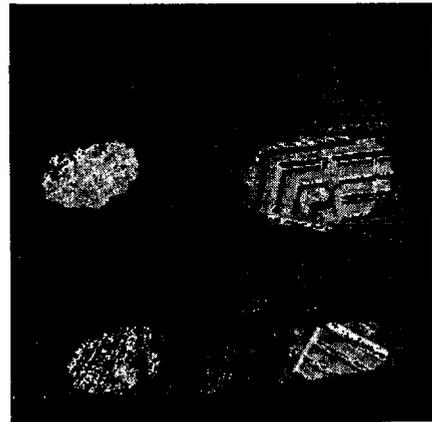
To demonstrate some of those issues, let us consider the “ $J=4$ ” simulated pair manifested in Fig. 3.4(c,d). The resulting AID between the overlapping areas of the aligned images yields PSNR=28.88 dB using the LS-ASLIV_{6,8} approach as shown in Fig. 3.5(d). The AID reflects how well the two overlapping areas of the aligned images are correlated, thereby giving an impression of the geometric registration precision and to what extent the illumination variations are corrected. Black regions refer to no difference in intensity, while non-zero differences show some level of mismatch between the aligned images due to either a sub-pixel geometric misalignment or a lack of perfect illumination correction. Let us, then, focus on the bottom-left quadrant depicted in Fig. 3.5(d). In this quadrant, we can make a number of observations. First, we see that remnants of the bottom shadow from Fig. 3.4(d) still exists. Throughout the shadowed region, the AID is larger, on average, than outside of the shadowed region. To exhibit the reason of this mismatch, we compare the estimated illumination parameters of this certain observed region, gain = 0.476 and offset = 14.846, to the ground truth values 0.5 and 8.73, respectively. Also, one can notice that there is still mis-estimation in the geometric parameters as shown in Table 3.1. We also note that a bright “halo” exists around the shadowed region’s boundary, showing a larger difference in the images. Generally, one can see small and large residuals in that AID. We observe

the same issue in the other AIDs of the resulting aligned pairs.

We can analyze the previous issue as follows. The illumination variations in all J regions do affect the behavior of the quadratic cost function in (3.9). The cost function squares the residuals of the images to be aligned, at each iteration until convergence occurs. So, if we suppose that the initial error of a certain region is large, then the resulting sum of squared errors of this region will be larger. As well, the LSE in (3.9) is expected to be sensitive to outliers. Thus, the resulting estimate of Φ is expected to be imprecise compared to what it should be (*i.e.*, the geometric and illumination parameters are negatively affected). We can argue that the main reason for such an issue is that the current estimator squares the whole residuals. To overcome this problem, we propose replacing the current estimator with a robust one. The next chapter addresses that problem in detail to improve the quality of the estimates.



(a) LS-GIM; PSNR=20.47 dB



(b) LS-AIM; PSNR=22.09 dB

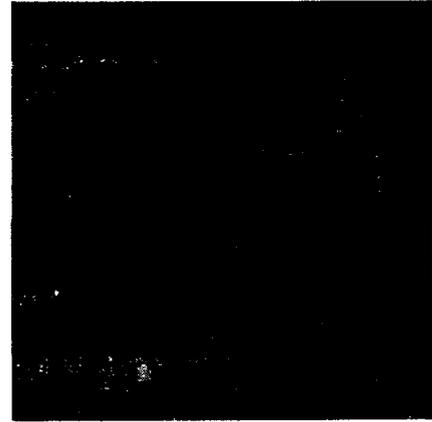
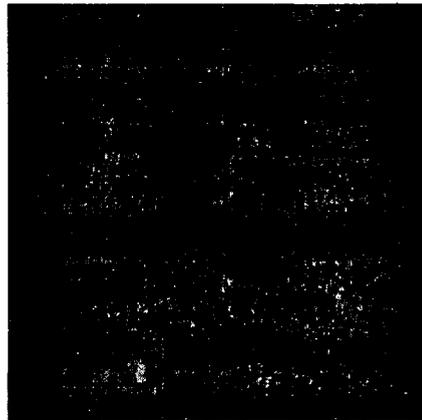
(c) LS-ASLIV_{6,6}; PSNR=27.69 dB(d) LS-ASLIV_{6,8}; PSNR=28.88 dB(e) LS-ASLIV_{6,10}; PSNR=28.80 dB

Figure 3.5: Normalized AIDs of the aligned images for the “ $J=4$ ” simulated pair manifested in Fig. 3.4(c,d), pair #45, using competing approaches.



(a) LS-GIM; NCC=0.8819



(b) LS-AIM; NCC=0.8998

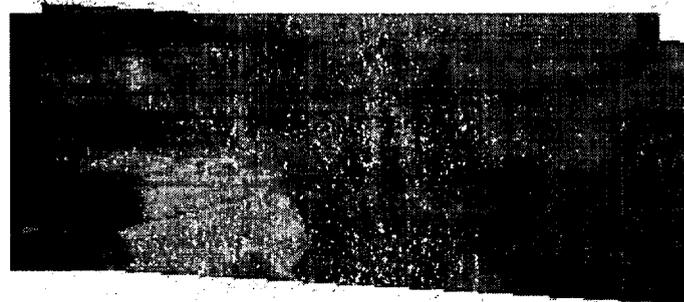
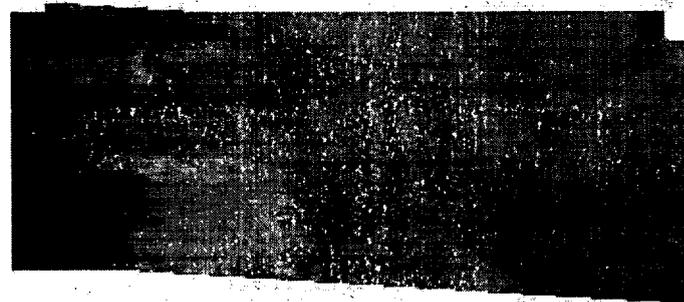
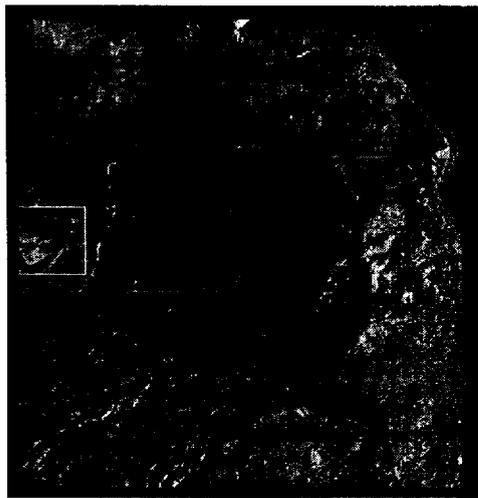
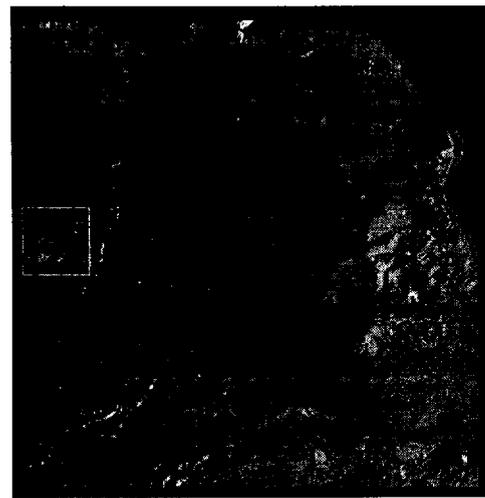
(c) LS-ASLIV_{6,6}; NCC=0.9226(d) LS-ASLIV_{6,8}; NCC=0.9384

Figure 3.6: Image mosaics of the natural pair shown in Fig. 2.1(a,b). Recall that the higher the NCC value, the better the performance of the approach.



(a) LS-GIM; NCC=0.9642



(b) LS-AIM; NCC=0.9691

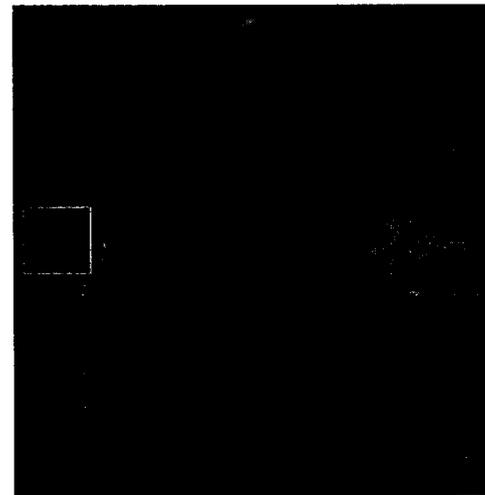
(c) LS-ASLIV_{6,6}; NCC=0.9789(d) LS-ASLIV_{6,8}; NCC=0.9800

Figure 3.7: Gamma-corrected AIDs of the real pair #B depicted in Fig. 2.1(c,d), $\gamma=1.7$. The darker the AID, the better the approach performance.

Chapter 4

Image Registration Using The M -estimator

4.1 Introduction

In the previous chapter, we presented the LS-ASLIV approach that exploits the proposed registration model, ASLIV, with the LSE that is sensitive to outliers. Thus, some areas of the AID of the aligned image pair have larger residuals that could be reduced. The challenge then lies in replacing the traditional LSE by one of the robust estimators that is not as susceptible as LSE to outliers, such as M -estimation presented by Huber [16].

In this chapter, we present an approach to overcome the drawbacks of the previous approach, LS-ASLIV. In Section 4.2, we propose a general framework of M -estimation jointly with the ASLIV registration model. In Section 4.3, we provide an iterative M -estimator scheme using the Huber function. In Section 4.5, the experiments are developed on real and simulated data sets using competing approaches. Finally, the M -estimator approach is analyzed in Section 4.6.

4.2 A General Framework of M -estimation With the ASLIV Registration Model

In this section, we introduce a registration approach using a robust M -estimator [16] jointly with the ASLIV registration model that is presented in Section 3.3. The approach is generic enough to employ any selected objective function.

Recall that the LSE issue is that it squares both small and large residuals for each illumination region. Instead, we propose using a robust M -estimator whose general form [16] can be cast as

$$\min_{\Omega} \left\{ \mathcal{L} \equiv \sum_{i=1}^n \rho(e_i) \right\}, \quad (4.1)$$

where Ω denotes the unknown vector to be estimated, e_i refers to the i^{th} residual of n observations, and ρ is a reasonable objective function that should have certain properties. Following (4.1), the minimization problem in (3.9) can be then rewritten as

$$\min_{\Phi} \left\{ \mathcal{L} \equiv \sum_{\mathbf{x}} \rho(E(\Phi)) \right\}, \quad (4.2)$$

where the residuals $E(\Phi)$ can still be approximated by (3.14). It is useful to note that M -estimation becomes LSE if $\rho(\cdot)$ in (4.1) is chosen to be a quadratic function (*i.e.*, the LSE is a special case of the M -estimator). There is still one issue to complete the mathematical derivation to estimate Φ , that is, choosing the objective function, ρ .

Much research exists in how to choose the cost function, ρ , in the M -estimator in order to obtain more accurate estimates. For example, Li [108] has shown that ρ should satisfy some aspects, such as

- Positivity: $\rho(t) \geq 0 \forall t$, with $\rho(0) = 0$,
- Similarity: $\rho(t) = \rho(-t)$,
- Convex Shape: $\rho(|t + 1|) \geq \rho(|t|) \forall t$, and
- Differentiability: ρ is continuously differentiable.

Let $\psi = \rho'$ be the derivative of ρ . Differentiating the objective function w.r.t. the coefficients, Φ , and setting the partial derivatives to 0, produces a system of equations to estimate Φ . It can be noticed that solving the estimation equations depends on the residuals. Also, the residuals rely upon the estimated coefficients. An iterative scheme is, therefore, required. The next section presents an iterative framework using M -estimation with a chosen cost function.

4.3 An Iterative Scheme Using the M -estimation With the ASLIV Model

In this section, we present a registration approach using M -estimation in an iterative coarse-to-fine scheme.

Before going through the registration approach, we should first choose an objective function, $\rho(t)$. The characteristics aforementioned could be satisfied by many potential functions [16, 17, 109–114] as reported in Table 4.1 with the corresponding first derivative functions, $\psi(t)$. Table 4.1 shows some candidate cost functions that are able to differently penalize the resulting small and large residuals. For instance, the behavior of $\rho_{HL}(t)$ varies in accordance with whether the value of t is small or large w.r.t. the dynamic range of the input images [0,255]. Thus, such cost functions do not specify whether the value of t is smaller or larger

Table 4.1: Six examples of potential functions $\rho(\cdot)$ that satisfy certain properties of M -estimator, with their associated first derivative functions $\psi(\cdot)$.

Potential function	$\rho(t)$	$\psi(t)/2$	References
$\rho_{TK}(t)$	$\lambda^2 t^2$	$\lambda^2 t$	[109]
$\rho_{GM}(t)$	$\frac{t^2}{1+t^2}$	$\frac{t}{(1+t^2)^2}$	[110]
$\rho_{GR}(t)$	$2 \log[\cosh(t)]$	$\begin{cases} t, & t = 0 \\ \tanh(t), & t \neq 0 \end{cases}$	[111]
$\rho_{HL}(t)$	$\log(1+t^2)$	$\frac{t}{1+t^2}$	[112]
$\rho_{HS}(t)$	$2\sqrt{1+t^2} - 2$	$\frac{t}{\sqrt{1+t^2}}$	[114]
$\rho_{HB}(t; \alpha)$	$\begin{cases} t^2, & t \leq \alpha, \\ 2\alpha t - \alpha^2, & t > \alpha \end{cases}$	$\begin{cases} -\alpha, & t < -\alpha, \\ t, & -\alpha \leq t \leq \alpha, \\ \alpha, & t > \alpha \end{cases}$	[16, 17]

w.r.t. the residual's range. Therefore, we propose using a cost function whose behavior changes due to a tuning parameter. For instance, the Tikhonov function $\rho_{TK}(t)$, listed in Table 4.1, is controlled by a tuning parameter, λ . However, the Tikhonov function and such similar functions, which have the same expression along with both small and large residuals, do not allow for different penalties to the residuals. Therefore, we will use the Huber function $\rho_{HB}(t; \alpha)$ as the cost function $\rho(\cdot)$ in (4.2), where α is an adaptive positive threshold by which the function differently behaves.

It is worth noting that for small residuals (*i.e.*, $|t| \leq \alpha$), the Huber function $\rho(\cdot)$ reduces to the usual quadratic function, while the Huber function becomes an l_1 penalty function for large residuals (*i.e.*, $|t| > \alpha$). As well, the Huber influence function is bounded as the residuals increase, thereby ultimately improving the

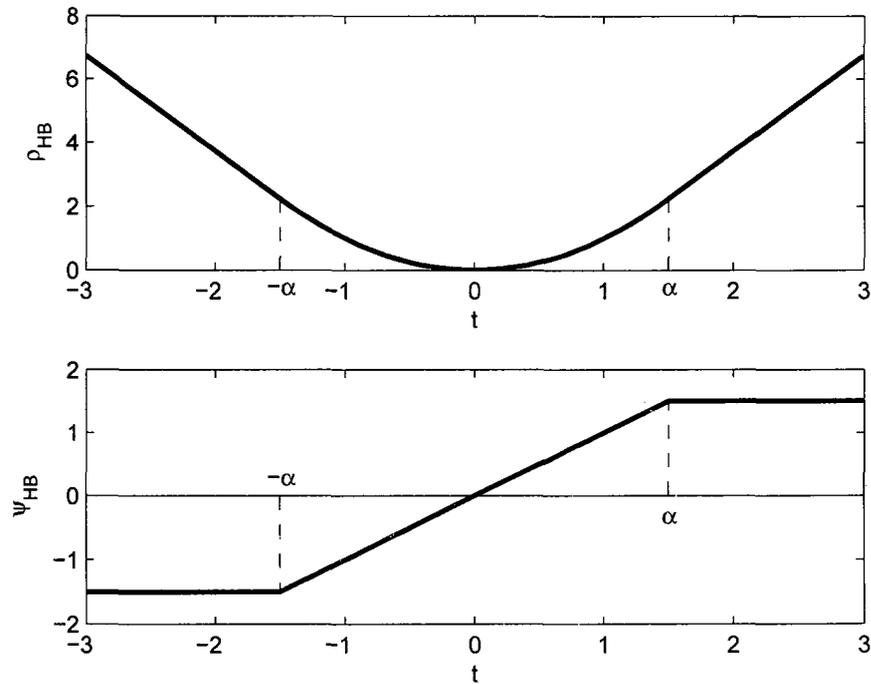


Figure 4.1: ρ_{HB} is the Huber function, and ψ_{HB} is its influence function [115].

estimates [16, 17]. Hence, we can reformulate (4.2) using the Huber function as

$$\min_{\Phi} \left\{ \mathcal{L} \equiv \sum_{\mathbf{x}} \rho_{HB}(E(\Phi); \alpha) \right\}. \quad (4.3)$$

Many methods have been developed to solve the minimization problem in (4.3). Some Newton type methods have been proposed in [116] leading to a possible nonconvergence because the objective function in (4.3) is not twice differentiable. Other reasonable complex Newton type methods are considered in [117–119]. In [120], a minmax formulation yielding a mixed complementarity problem was suggested, which afterwards reduced to a quadratic formulation as in [121].

However, Mangasarian and Musicant [121] showed that the Huber M -estimator, $\forall \alpha > \bar{\alpha}$ for some $\bar{\alpha}$, could be cast as a minimum of a convex quadratic function. Mangasarian and Musicant also exhibited that their proposed method is considerably faster than others. Following [121], (4.3) can be rewritten as

$$\min_{\Phi} \left\{ \mathcal{L} = 0.5 \sum_{\mathbf{x}} [F \cdot E^2(\Phi) + 2\alpha S \cdot E(\Phi) - \alpha^2] \right\}, \quad (4.4)$$

where (\cdot) refers to element-by-element matrix product. The approach in [121] is based on separating the small residuals from the large ones to quadratic and linear penalties, respectively. The separation step can be performed using two matrices: (i) the large-residual selective matrix which is expressed as

$$S = \begin{bmatrix} s_{11} & \dots & s_{1M} \\ \vdots & \ddots & \vdots \\ s_{N1} & \dots & s_{NM} \end{bmatrix}, \quad (4.5)$$

where

$$s_{xy} = \begin{cases} -1, & E(\Phi; \mathbf{x}) < -\alpha \\ 0, & |E(\Phi; \mathbf{x})| \leq \alpha \\ 1, & E(\Phi; \mathbf{x}) > \alpha \end{cases}, \quad (4.6)$$

and (ii) the small-residual selective matrix that can be represented as

$$F = \mathbf{1} - S \cdot S, \quad f_{xy} = 1 - s_{xy}^2, \quad (4.7)$$

where f_{xy} is an element of F at position (x, y) , and $\{\mathbf{1}, S, F\} \in \mathbb{R}^{N \times M}$. Note that $E(\Phi; \mathbf{x})$ is an element located at position (x, y) in matrix $E(\Phi)$.

Following Section 3.4, the unknown vector Φ can be estimated using the

Gauss-Newton algorithm [101] to solve the non-linear minimization problem in (4.4). Recall that $\hat{\Phi}$ is iteratively updated by (3.17) at each iteration g . Also, recall that the residuals, $E(\Phi)$, can be still represented by (3.14) assuming the affine motion model in (3.15). Replacing $E(\cdot)$ in (4.4) by its 1st order Taylor series expansion, then \mathcal{L} in (4.4) can be rewritten as

$$\mathcal{L} = 0.5 \sum_{\mathbf{x}} \left[F \cdot (E(\Phi_{g-1}) + \Delta_g^\top \nabla_{\Phi} E(\Phi_{g-1}))^2 + 2\alpha S \cdot (E(\Phi_{g-1}) + \Delta_g^\top \nabla_{\Phi} E(\Phi_{g-1})) - \alpha^2 \right]. \quad (4.8)$$

Setting the gradient of \mathcal{L} in (4.8) w.r.t. Δ to zero, we obtain

$$-\sum_{\mathbf{x}} [(E(\Phi_{g-1}) \nabla_{\Phi} E(\Phi_{g-1})) \cdot F + \alpha S \cdot \nabla_{\Phi} E(\Phi_{g-1})] = \hat{\Delta}_g^\top \sum_{\mathbf{x}} [(\nabla_{\Phi} E(\Phi_{g-1}) \nabla_{\Phi} E^\top(\Phi_{g-1})) \cdot F]. \quad (4.9)$$

We can write (4.9) in a matrix notation as

$$-\mathbf{Y} \mathbf{P}^\top = (\mathbf{Y} \mathbf{H}^\top) \hat{\Delta}, \quad (4.10)$$

where

$$\mathbf{H} = [H_{1,1}, H_{1,2}, \dots, H_{N,M}], \quad (4.11)$$

$$\mathbf{Y} = [Y_{1,1}, Y_{1,2}, \dots, Y_{N,M}], \quad (4.12)$$

$$\mathbf{P} = [P_{1,1}, P_{1,2}, \dots, P_{N,M}], \quad (4.13)$$

$$H_{nm} = [n I_x, m I_x, n I_y, m I_y, I_x, I_y, -I_1 Q_1, \dots, -I_1 Q_J, -Q_1, \dots, -Q_J]^\top, \quad (4.14)$$

$$Y_{nm} = f_{nm} [n I_x, m I_x, n I_y, m I_y, I_x, I_y, -I_1 Q_1, \dots, -I_1 Q_J, -Q_1, \dots, -Q_J]^T, \quad (4.15)$$

and

$$P_{nm} = \alpha s_{nm} + f_{nm} I_2(n, m). \quad (4.16)$$

It can be shown that equations (4.10) through (4.16) can be used to perform one iteration for finding a solution of $\hat{\Delta}$. The unknown vector, Φ , is iteratively accumulated due to (3.17) yielding a single final estimate.

In this section, a registration approach using the Huber M -estimator is presented in order to eventually improve the GRA jointly with illumination correction. It can be shown that this approach is independent of the cost function as long as the latter satisfies certain conditions. We will refer to this Huber M -estimator approach explained in this chapter, by HM-ASLIV_{6,2J}, where the superscript accounts for the $6+2J$ parameters in Φ . Note that the HM-ASLIV approach still exploits the iterative coarse-to-fine framework and stopping criteria as mentioned in Section 3.4 and Section 3.7, respectively.

4.4 Summary on the HM-ASLIV Approach

Thus far, we presented the registration approach, HM-ASLIV, using a robust M -estimator with the Huber function. Algorithm 4.1 summarizes the proposed approach HM-ASLIV that is cast in an iterative coarse-to-fine framework.

Algorithm 4.1 The proposed image registration approach HM-ASLIV

-
- 1: **Given:** two input images, I_1 and I_2 , as described in Section 3.3.
 - 2: **Required:** vector, Φ , that includes the finally estimated geometric and illumination parameters.
 - 3: **Initialize:**
 - 4: Φ ... the unknown vector by Φ° ,
 - 5: r ... the no. of resolution levels,
 - 6: g ... the max no. iterations per resolution level,
 - 7: ε ... a predefined threshold of the cost function update,
 - 8: J ... the no. of illumination regions, and
 - 9: i ... iteration counter by 1.
 - 10: **for** $p = 1$ to r **do**
 - 11: **repeat**
 - 12: Apply $\Phi^{(i)}$ to I_1 .
 - 13: Determine the AID⁽ⁱ⁾: $I_{21} = |I_2 - I_1^{\Phi^{(i)}}|$.
 - 14: Apply the k -means algorithm to the AID⁽ⁱ⁾ to obtain the binary masks, $Q_j^{(i)}$, of the illumination regions, R_j° , setting k to J .
 - 15: Determine the residuals, $E(\Phi^{(i)})$, using (3.14).
 - 16: Determine the standard deviation, $\sigma^{(i)}$, of the current residuals.
 - 17: Set the cost function's (*i.e.*, the Huber function) threshold, $\alpha^{(i)}$, to $1.345 \sigma^{(i)}$ [17, 101].
 - 18: Determine the large-residual selective matrix, $S^{(i)}$, using (4.5) and (4.6).
 - 19: Determine the small-residual selective matrix, $F^{(i)}$, using (4.7).
 - 20: Determine the corresponding value of the cost function, $\mathcal{L}(\Phi^{(i)})$, using (4.8).
 - 21: Determine the cost function update $\delta_i = \mathcal{L}(\Phi^{(i)}) - \mathcal{L}(\Phi^{(i-1)})$.
 - 22: Determine the unknown vector's update, $\Delta_{(i)}$, using (4.10).
 - 23: Update $\Phi^{(i)}$ by $\Delta_{(i)}$ using (3.17).
 - 24: $i \leftarrow i + 1$.
 - 25: **until** ($\delta_{(i)} < \varepsilon$) or ($i > g$)
 - 26: **end for**
-

4.5 Experiments & Results

Before we demonstrate the experiments, first, we would like to emphasize that we use the same real and simulated data sets as described in Section 3.5. Also, the implementation parameters are set as mentioned in Section 3.7, (*i.e.*, the stopping threshold $\varepsilon = 0.1$, the maximum number of iterations $g = 10$). As well, Φ° is initialized either by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0]$ using the ASLIV_{6,6}-based approaches or by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0]$ with the ASLIV_{6,8}-based approaches.

In the HM-ASLIV approach, there are many ways to choose the threshold, α , of Huber function. One suggestion [122] is to set α as $[1.5\sigma, 2\sigma]$, where σ is the standard deviation of the residuals at each iteration. Another suggestion [123] is to let $\alpha = \text{MAR}/0.6745$, where MAR is the median absolute residual. We set α to 1.345σ following [17, 101] to eliminate 5% of the outliers assuming that the residuals are normally distributed.

In this section, the experiments are performed with the HM-ASLIV approach compared to the LS-GIM [10, 11], LS-AIM [13] and the LS-ASLIV [18, 19] approaches using the same performance evaluation measures stated in Section 3.6. Recall that the LS-GIM implementation can be found online at [107] provided by Aguiar [11].

4.5.1 Experiments on Simulated Data Sets

The first experiment with HM-ASLIV is performed on both “ $J=3$ ” and “ $J=4$ ” data set using the LS-GIM, LS-AIM, LS-ASLIV, HM-ASLIV approaches considering the ASLIV version (*i.e.*, ASLIV_{6,6}-based approaches uses $J=3$, while ASLIV_{6,8}-based approaches employs $J=4$).

Table 4.2: The AAEs ($\times 10^{-4}$) of the estimated affine parameters of the “ $J=3$ ” data set. The less the value of AAEs, the better the performance of the approach.

	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}
a_1	15.3658	12.1419	1.7205	1.2025
a_2	12.0017	10.5204	1.5079	1.0344
a_3	14.6081	11.9571	1.6189	1.1661
a_4	19.1128	15.8924	2.0502	1.8699
a_5	9190.76	7649.53	789.16	722.35
a_6	2688.72	2236.18	795.70	716.11

Table 4.3: The AAEs ($\times 10^{-4}$) of the estimated affine parameters of the “ $J=4$ ” data set. The less the value of AAEs, the better the performance of the approach.

	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}
a_1	16.5962	13.7167	1.8121	1.4358
a_2	13.2015	10.6395	1.6025	1.1916
a_3	15.5246	13.2940	1.7360	1.2049
a_4	21.0259	17.3674	2.4261	1.9280
a_5	9472.09	7918.66	816.50	760.57
a_6	2835.12	2382.50	804.81	755.11

Table 4.2 and Table 4.3 show the AAEs of the estimated affine parameters of the “ $J=3$ ” and “ $J=4$ ” data sets, respectively, computed by the LS-GIM, LS-AIM, LS-ASLIV and HM-ASLIV approaches. The HM-ASLIV approach provides more precise motion estimates, in the presence of local illumination variations, by an average increase of 88.5%, 82.3% and 18.9% more than the LS-GIM, LS-AIM and LS-ASLIV approaches, respectively. The scalar threshold could segment the residuals into small and large areas that are differently penalized using the Huber function. Thus, the final estimates are refined leading to an improved GRA using the robust M -estimator.

Table 4.4 shows the effectiveness of the HM-ASLIV approach in terms of SSIM indexes and PSNR values, using the LS-GIM, LS-AIM, LS-ASLIV_{6,6} and HM-ASLIV_{6,6} approaches with the “ $J=3$ ” data set. It can be noticed that the HM-ASLIV

approach outperforms the LS-GIM, LS-AIM and LS-ASLIV approaches by an average increase of 14.8%, 11.6% and 3.5% in SSIM and by an average increase of 27.1%, 24.8% and 5.8% in PSNR, with a slight average increase of 33.1% 29.2% and 11.6% in computational time, respectively, as shown in Table 4.5.

Similarly, Table 4.6 shows the SSIM indexes and PSNR values using the LS-GIM, LS-ASLIV_{6,8} and HM-ASLIV_{6,8} approaches with “ $J=4$ ” data set. One can notice that the HM-ASLIV approach outperforms the LS-GIM, LS-AIM and LS-ASLIV approaches by an average increase of 12.5%, 10.1% and 3.1% in SSIM and by an average increase of 26.3%, 23.9% and 5.2% in PSNR, with a slight average increase of 33.8%, 30.1% and 12.2% in computational time, respectively, as shown in Table 4.7.

In Fig. 4.2(a,b,c), we show the normalized AIDs of the aligned pairs of the “ $J=3$ ” simulated pair shown in Fig. 3.4(a,b) using the LS-GIM, LS-ASLIV_{6,6} and HM-ASLIV_{6,6} approaches, respectively. Depicted is that the HM-ASLIV approach, in Fig. 4.2(c) with PSNR=27.76 dB, surpasses those in Fig. 4.2(a) and (b) using the LS-GIM and LS-ASLIV approaches with PSNR=21.91 and 27.01 dB, respectively. As well, one can notice that the two bottom illumination regions are brighter in Fig. 4.2(b) using LS-ASLIV approach as opposed to Fig. 4.2(c) using HM-ASLIV approach. As well, one can see that halos in Fig. 4.2(b) using the LS-ASLIV_{6,6} has nearly disappeared in Fig. 4.2(c) using the HM-ASLIV_{6,6}.

Table 4.4: SSIM indexes and PSNR values for a random subset of the “ $J=3$ ” data set using the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6} and (iv) HM-ASLIV_{6,6} approaches.

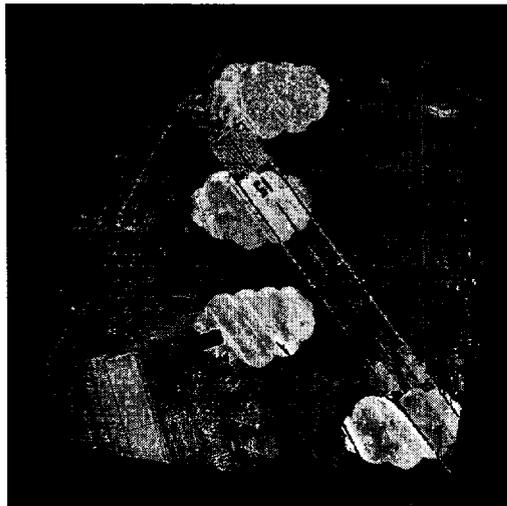
Pair #	SSIM [104]				PSNR (dB)			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
10	0.8397	0.8530	0.9602	0.9622	22.53	24.79	29.02	29.78
17	0.8194	0.8379	0.9577	0.9589	20.01	22.14	28.33	29.01
28	0.8268	0.8645	0.9534	0.9598	21.76	24.12	29.42	30.17
34	0.8066	0.8358	0.9507	0.9531	21.91	23.35	27.01	27.76
38	0.8552	0.8853	0.9650	0.9665	23.47	24.93	28.63	29.25
42	0.8201	0.8413	0.9612	0.9681	22.95	24.89	29.25	29.87
48	0.8419	0.8742	0.9634	0.9653	21.94	22.05	29.69	30.10

Table 4.5: Computational time (sec.) for a random subset of the “ $J=3$ ” data set with the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6} and (iv) HM-ASLIV_{6,6} approaches.

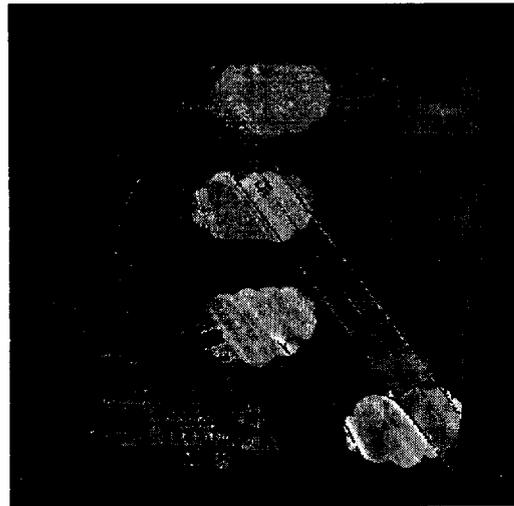
#	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}
10	43.0	44.6	57.7	62.2
17	42.8	43.8	55.4	59.8
28	44.7	46.0	58.8	61.4
34	43.3	44.9	57.2	62.5
38	42.7	44.1	56.4	63.2
42	43.4	45.2	56.5	60.0
48	41.2	43.4	55.3	59.4

Table 4.6: SSIM indexes and PSNR values for a random subset of the “ $J=4$ ” data set using the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,8} and (iv) HM-ASLIV_{6,8} approaches.

Pair #	SSIM [104]				PSNR (dB)			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
10	0.7969	0.8295	0.9341	0.9417	22.12	23.97	28.57	29.15
17	0.7844	0.8174	0.9254	0.9363	19.71	21.75	27.66	28.35
28	0.8021	0.8348	0.9164	0.9227	21.37	23.42	28.95	29.76
34	0.7721	0.8039	0.9142	0.9253	21.51	23.36	26.66	27.46
38	0.8441	0.8762	0.9284	0.9494	22.99	24.83	27.93	28.77
42	0.7778	0.8103	0.9334	0.9405	22.60	24.75	28.62	29.39
48	0.7999	0.8329	0.9245	0.9306	21.67	23.81	29.18	29.82



(a) LS-GIM, PSNR=21.91 dB



(b) LS-AIM, PSNR=23.35 dB

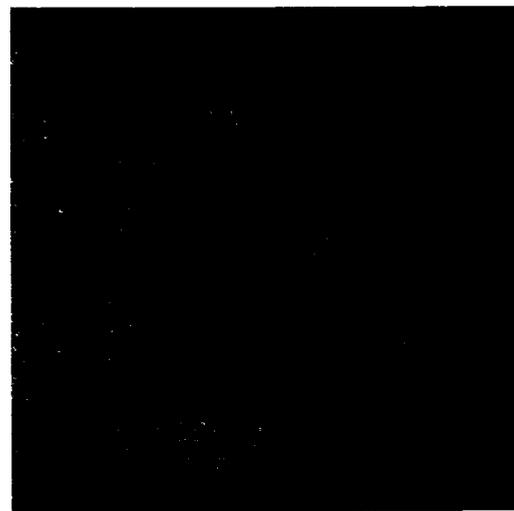
(c) LS-ASLIV_{6,6}, PSNR=27.01 dB(d) HM-ASLIV_{6,6}, PSNR=27.76 dB

Figure 4.2: Using competing approaches, (a,b,c) normalized AIDs of the “ $J=3$ ” simulated pair #34 depicted in Fig. 3.4(a,b). The darker the AID, the better the approach performance.

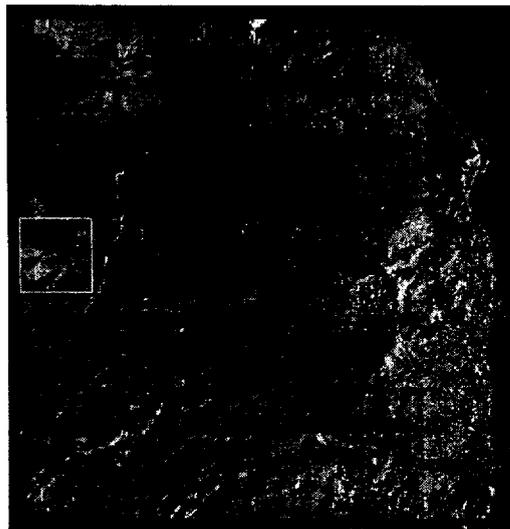
Table 4.7: Computational time (sec.) for a random subset of the “ $J=4$ ” data set with the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,8} and (iv) HM-ASLIV_{6,8} approaches.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}
10	43.4	45.8	59.0	63.7
17	44.0	47.3	57.3	60.9
28	46.6	49.3	60.2	62.8
34	45.4	48.1	58.8	63.8
38	44.5	47.8	58.0	65.1
42	43.9	46.5	57.7	61.1
48	42.9	45.4	56.8	60.7

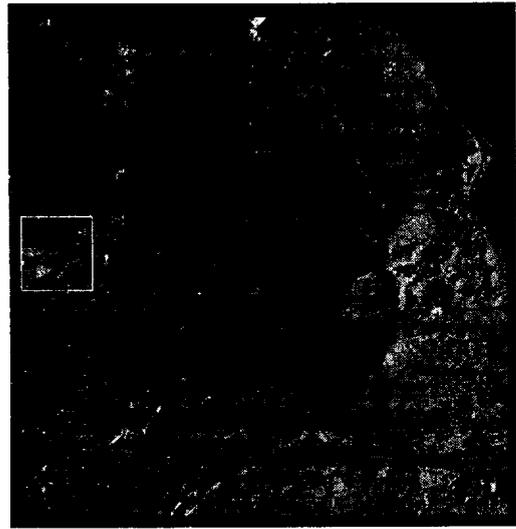
4.5.2 Experiments on the Real Data Set

This section describes the experiments that were performed on the real data set as described in Section 3.5 using LS-GIM, LS-AIM, LS-ASLIV_{6,6} and HM-ASLIV_{6,6} approaches with the k -means algorithm, with $k=J=3$.

Table 4.8 shows resulting NCC values that express the correlation between the overlapping areas of the two aligned images using competing approaches. It is clear that the HM-ASLIV approach surpasses the LS-GIM, LS-AIM and LS-ASLIV approaches in terms of NCC. To further illustrate, Fig. 4.3(a,b,c,d) shows gamma-corrected AIDs of the registered pairs, of the real pair in Fig. 2.1(c,d) using LS-GIM, LS-AIM, LS-ASLIV_{6,6} and HM-ASLIV_{6,6} approaches, with NCC = 0.9642, 0.9691, 0.9789 and 0.9802, respectively ($\gamma = 1.7$ for better visualization). Fig. 4.3(d) has many areas that are darker than the corresponding areas in Fig. 4.3(a), (b) and (c). Blown up views of the white boxes in Fig. 4.3(a,b,c,d) are shown in Fig. 4.4(a,b,c,d), respectively, depicting the same observation in Fig. 4.3.



(a) LS-GIM; NCC=0.9642



(b) LS-AIM; NCC=0.9691

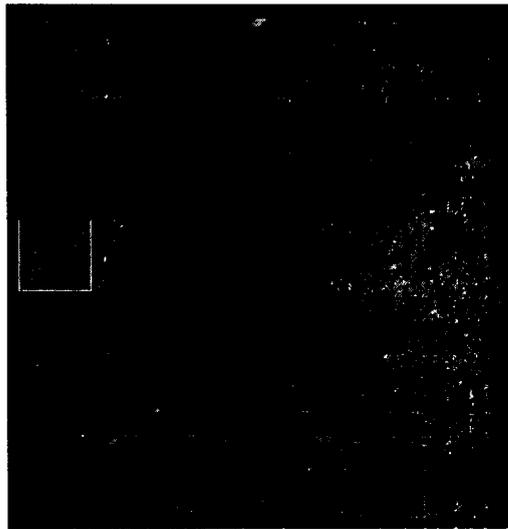
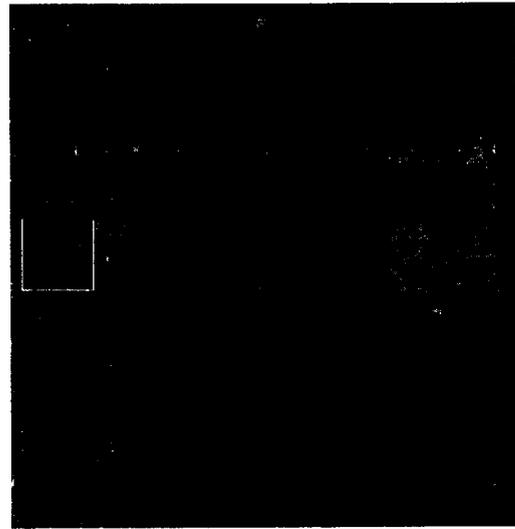
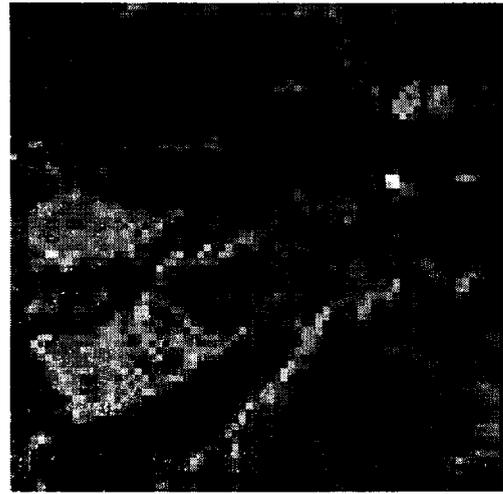
(c) LS-ASLIV_{6,6}, NCC=0.9789(d) HM-ASLIV_{6,6}, NCC=0.9802

Figure 4.3: The gamma-corrected AIDs for the real pair in Fig. 2.1(c,d), pair #B, $\gamma = 1.7$, $J=3$. The darker the AID, the better the approach performance.



(a) LS-GIM



(b) LS-AIM

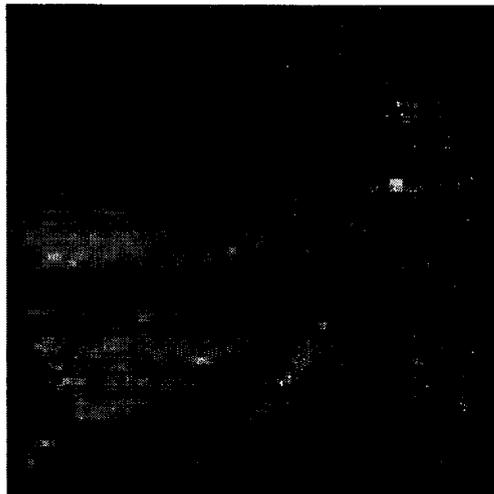
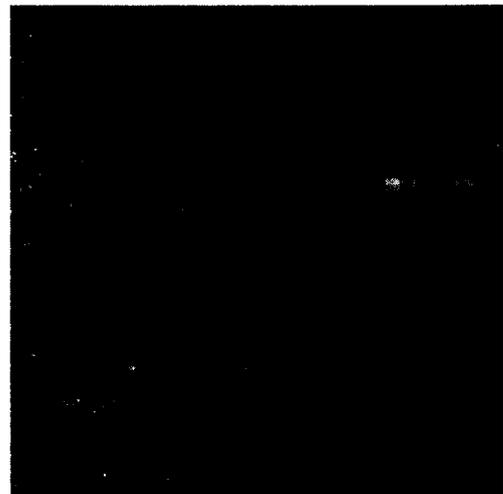
(c) LS-ASLIV_{6,6}(d) HM-ASLIV_{6,6}

Figure 4.4: Blow-up views of the white boxes in Fig. 4.3, $\gamma = 1.7$. The darker the AID, the better the approach performance.

Table 4.8: NCC values for real data set using the (i) LS-GIM, (ii) LS-AIM, (iii) LS-ASLIV_{6,6} and (iv) HM-ASLIV_{6,6} approaches. Note that pair #A and #B refer to Fig. 2.1(a,b) and (c,d), respectively.

Pair	(i)	(ii)	(iii)	(iv)
A	0.8819	0.8998	0.9226	0.9237
B	0.9642	0.9691	0.9789	0.9802
C	0.9799	0.9813	0.9853	0.9870
D	0.9450	0.9520	0.9644	0.9659
E	0.9601	0.9665	0.9743	0.9758
F	0.9719	0.9786	0.9922	0.9939
G	0.9551	0.9594	0.9657	0.9671
H	0.8971	0.9087	0.9229	0.9245
I	0.9081	0.9105	0.9241	0.9286
J	0.9542	0.9604	0.9770	0.9833

4.6 Analysis & Discussion

In this section, we analyze the HM-ASLIV approach that is shown in this chapter and presented in [20]. We replace the LSE with a more robust estimator, such as an M -estimator that exhibits more resistance to outliers. For this, the Huber function is exploited as the cost function. The robustness of the Huber function comes from its nature in that it linearly and quadratically penalizes large and small residuals, respectively. This robustness to outliers yields more precise estimates than using the LSE. However, it can be shown, *e.g.*, in the white box of Fig. 4.3(c) that some areas still have large residuals compared to the surrounding areas. One can observe this problem throughout all the AIDs of aligned simulated and real images.

We can analyze the issue above as follows. The tuning threshold, α , in the cost function, ρ , is used for all of the residuals. In other words, let us focus on a certain illumination region. Some residuals of that specific region could be classified as

large w.r.t. all residuals using the underlying threshold. Whereas, the same residuals could be classified as small w.r.t. the residuals of that illumination region. Thus, larger residuals result because the chosen threshold may be reasonable for some illumination regions, but not for others. Note that this challenge still holds for any objective function with a single-threshold. Therefore, we believe each illumination region should have its own proper threshold, thereby adapting to the characteristics of the region. In addition, during the registration process, the boundary of each illumination region may contain some areas that do not belong to the right region. This segmentation error is due to iteratively using a clustering algorithm to create the illumination regions. These mis-segmented areas negatively impact the GRA.

In the following chapter, we propose an approach that uses different M -estimators. Each is supposed to have a distinct cost function with its own threshold. As well, a weighting function is provided to meet the negative effects of the segmentation errors.

Chapter 5

Image Registration Using Weighted Region-based M -estimators

5.1 Introduction

In Chapter 3, the LS-ASLIV approach is presented using LSE in conjunction with the ASLIV registration model that copes with images having variant illumination changes of arbitrary shapes [18, 19]. However, the influence function of LSE is proportional to the residuals, thus being very vulnerable to outliers. Then, in Chapter 4, we replace LSE with a more robust estimator, such as M -estimators [20]. Given a cost function having a threshold, the residuals have been segmented into two areas: small and large residuals. Then, each area is differently penalized using the nature of the cost function having a threshold. One problem in this approach is that the selected threshold is not necessarily appropriate for all illumination regions, thus yielding larger error in some areas. In addition, during the registration process, the boundary of each illumination region may contain some pixels that belong to its adjacent region. The challenges are how to create a locally spatially adaptive threshold for each region and how to reduce the negative

effects of such mis-segmented pixels to eventually improve the GRP.

In this chapter, we propose a registration approach using what we call weighted region-based M -estimators, such that each illumination region should have its own cost function in order to ultimately further improve the GRP. Fig. 5.1 exhibits the idea of the proposed approach. One can see that each illumination region, R_j , is segmented into large and small subregions, Λ_j and Γ_j , respectively. This segmentation is performed using a cost function with a distinct threshold, α_j , that corresponds to R_j .

The proposed approach is general enough to employ any objective function that satisfies the conditions in Section 4.2. For experimental work, we exploited a modified version of the Huber function as an objective function. We extended the Huber function from one threshold to multi-threshold in order to cope with J illumination regions. As well, the proposed approach could employ any bounded weighting function. For example, the Tukey function is used so as to reduce the negative impacts of the mis-segmented areas located on the boundary of each illumination region. The proposed approach is cast in an iterative coarse-to-fine scheme, not only to allow for large motion deformations, but also to reduce the computational cost and obtain accurate estimates.

The rest of this chapter is organized as follows. Section 5.2 proposes a registration approach using different M -estimators with a weighting for uncertainty in the segmentation to lessen the impact of mis-segmented pixels. Section 5.3 focuses on mathematically manipulating the proposed approach using M -estimators with a region-based cost function and a particular weighting function. Section 5.4 presents the implementation setup of the competing approaches. Section 5.5 summarizes the proposed approach. Section 5.6 exhibits

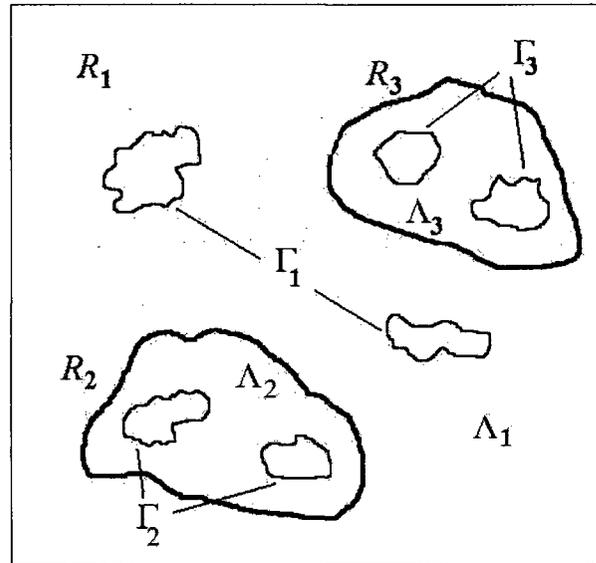


Figure 5.1: Each region, R_j , of the AID of two images to be aligned at a certain iteration, is segmented by α_j into small and large residual subregions, Λ_j and Γ_j , respectively.

the experiments developed on real and simulated data sets using competing approaches. Finally, analysis and discussion on the proposed approach are given in Section 5.7.

5.2 Weighted Region-Based M -estimators For Image Registration

In this section, we propose a new registration approach with different M -estimators in Section 5.2.1 and reduced-weighting for mis-segmented boundary areas in Section 5.2.2.

5.2.1 A General Framework of M -estimators With Region-based ASLIV Model

Since each region should have its own gain and offset as mentioned in (3.8), since LSE is vulnerable to outliers, and since a cost function with a scalar threshold may be unfair for all regions, generally speaking, we believe using a distinct M -estimator for each illumination region would improve the GRP and obtain illumination correction. In other words, we propose a registration approach that uses the ASLIV model in a framework of J different M -estimators, such that a distinct objective function, ρ_j , is assigned to each illumination region. Then, the minimization problem in (4.2) can be reformulated as

$$\min_{\Phi} \left\{ \mathcal{L} \equiv \sum_{j=1}^J \sum_{\mathbf{x} \in R_j} \rho_j(E_j(\Phi, \mathbf{x})) \right\}, \quad (5.1)$$

and the residuals of each region can be expressed as

$$E_j(\Phi) = Q_j \cdot E(\Phi), \quad (5.2)$$

where the residuals $E(\Phi)$ can still be approximated by (3.14) and $\{E_j\} \in \mathbb{R}^{N,M}$. We can see from (5.1) that each region is now handled with its own M -estimator, ρ_j .

5.2.2 A Weighting Function For Segmentation Deviations

This subsection addresses another issue, that is segmentation deviations, and proposes a solution.

As stated in Section 3.3, a segmentation algorithm is used to segment the AID

of the images to be aligned, yielding a binary mask, Q_j , for each illumination region, R_j . This mask is iteratively changed (refined) due to either applying the currently estimated Φ_k at iteration k , or upscaling the mask to another resolution level as shown in [18]. During this iterative process in a coarse-to-fine framework, some of the boundary pixels of an illumination region may be mis-segmented and may actually belong to an adjacent region. Hence, we propose decreasing the weights of those boundary pixels with the hopes of improving the GRP.

To exhibit the idea, let us focus on Fig. 5.2. The left part of the figure shows two masked illumination regions, R_1 and R_2 . The right part of the figure shows an enlarged view of the box showing boundary layers, D_1 and D_2 , respectively, for each region. Let's assume that each layer, $D_j \forall j \in \{1, 2\}$, consists of T rings $G_{j,t} \forall 1 \leq t \leq T$. We assume the boundary layer thickness is T pixels, while a boundary ring thickness is one pixel. In the right part of Fig. 5.2, the dash-dot-dot line shows the *a priori* known segmentation line, while the solid line depicts the computed segmentation line using a clustering algorithm. Should these two lines not match, then both regions, R_1 and R_2 , definitely have mis-segmented pixels (*e.g.*, see the grey-shaded areas in Fig. 5.2). To lessen the impact of such pixels, each ring, $G_{j,t}$, is assigned a weight according to its position w.r.t. the first ring, $G_{j,1}$, using a weighting function, such that

$$f_{\psi(t)}(t, T) = \begin{cases} \psi(t) \in [0, 1), & t < T \\ 1, & t \geq T \end{cases} \quad (5.3)$$

where $\psi(t)$ is a monotonically non-decreasing function. Note that (5.3) is written from one illumination region perspective. Let us consider that a boundary layer,

D_j , of an illumination region, R_j , consists of T rings as

$$D_j = \bigcup_{t=1}^T G_{j,t}. \quad (5.4)$$

Then, a corresponding weighting matrix, W_j , using an uncertainty function, $f(t, T)$, can be created as

$$W_j = \begin{bmatrix} w_{11} & \dots & w_{1M} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & w_{NM} \end{bmatrix}, \quad w_{xy} = \begin{cases} f(t, T), & (x, y) \in G_{j,t} \\ 1, & (x, y) \in R_j \setminus D_j \\ 0, & \text{Otherwise} \end{cases}. \quad (5.5)$$

Thus, only (5.2) will be changed to

$$E_j(\Phi) = W_j \cdot Q_j \cdot E(\Phi), \quad (5.6)$$

where $W_j \in \mathbb{R}^{N,M}$. The minimization problem in (5.1) and all other related equations still hold without change using the motion model in (3.5).

We can summarize the proposed registration approach as follows. First, we suggest use a distinct robust estimator, such as an M -estimator, for each illumination region to cope with its own residuals to ultimately improve the GRP. Meanwhile, the proposed approach imposes weights in the segmentation using a weighting function to those pixels located on the boundary of each illumination region due to mis-segmentation. We'd like to emphasize that the proposed approach is general enough:

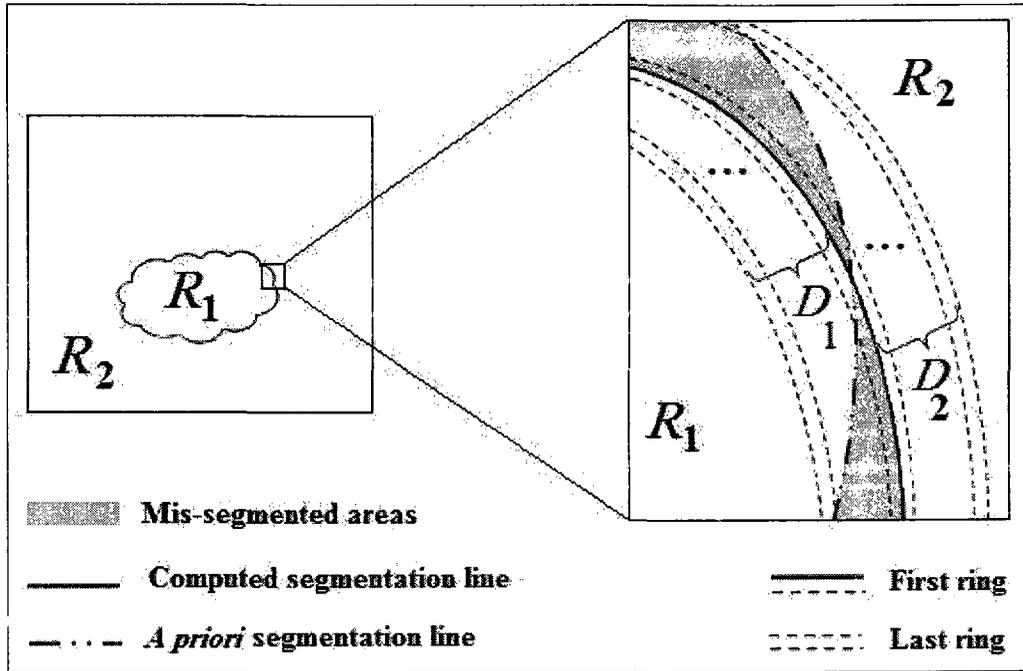


Figure 5.2: (Left) A masked image segmented into two illumination regions, $R_j, \forall j \in \{1, 2\}$. (Right) Blow up view of the box showing: the actual segmentation line, computed segmentation line, segmentation deviations (in grey), and first and last rings, $G_{j,1}$ and $G_{j,T}$, respectively, of both regions.

- To employ any objective function, ρ_j , provided that it meets some conditions, so that each illumination region could have its own objective function, and
- To exploit any weighting function that satisfies (5.3).

5.3 Proposed Approach Using M -estimator & Selected Functions

In the previous section, we have demonstrated the proposed approach in a general framework using an objective function with certain attributes in addition to

a weighting function and a generic motion model. In this section we specify a cost function, weighting function and a motion model to employ in the proposed approach. As well, we present a mathematical manipulation that allows us to estimate the unknown vector, Φ , through an iterative coarse-to-fine scheme.

5.3.1 M -estimation Using Huber & Tukey Functions

In this subsection, we mathematically present an iterative framework of the proposed approach to estimate the unknown vector, Φ .

First, we should select a weighting function following (5.3) to lessen the impact of the mis-segmented pixels located on the boundary layer of each illumination region. Recall that we assumed that each layer has a thickness of T pixels. Many functions could be used for weighting such boundary pixels, such as a delayed unit-step function, a sigmoid function, or the Tukey function. Since we do not focus on determining the optimal weighting function that should be employed in the proposed approach, we will choose the Tukey function [124] for $\psi(t)$ and rewrite (5.3) as

$$f(t, T) = \begin{cases} \frac{t^2}{T^2} - \frac{t^4}{T^4} + \frac{t^6}{T^6}, & |t| < T \\ 1, & \text{otherwise} \end{cases}, \quad (5.7)$$

where T is a tuning parameter by which the boundary layer thickness can be directly controlled.

Second, we have to choose another function to act as cost function, $\rho(t)$, as mentioned in Section 4.2. The properties of $\rho(t)$ can be achieved by many potential functions, such as Tikhonov [109], Geman-McClure [110], Green [111],

Hebert-Leahy [112], Rouchouze [113], Charbonnier [114], and Huber [16]. Generally speaking, all aforementioned functions and similar ones can be theoretically used by an M -estimator for each illumination region as proposed in Section 4.2.

We consider the Huber function

$$\rho(t; \alpha) = \begin{cases} \frac{1}{2}t^2, & |t| \leq \alpha \\ \alpha|t| - \frac{1}{2}\alpha^2, & |t| > \alpha \end{cases}, \quad (5.8)$$

since the other mentioned functions do not have a tunable threshold, α . Such a threshold adaptively controls the objective function's behavior to differently penalize small and large residuals, thereby fine tuning the estimates in (5.1). Using the Huber function, we believe each illumination region should have its own threshold. Since there are J regions, a certain threshold, α_j , should be assigned to an illumination region, R_j , (*i.e.*, each region should have its own Huber M -estimator).

In this chapter, we'd like to emphasize that we extend the ordinary formula of the Huber function in (5.8) from one threshold [16] to a region-based Huber function with multi-thresholds as

$$\rho(E_j(\Phi); \alpha_j) = \begin{cases} \frac{1}{2}E_j^2(\Phi), & |E_j(\Phi; \mathbf{x})| \leq \alpha_j, \forall \mathbf{x} \in R_j \\ \alpha_j|E_j(\Phi)| - \frac{1}{2}\alpha_j^2, & |E_j(\Phi; \mathbf{x})| > \alpha_j, \forall \mathbf{x} \in R_j \end{cases}, \quad (5.9)$$

where $E_j(\Phi; \mathbf{x})$ refers to an element in a matrix $E_j(\Phi)$ located at position \mathbf{x} . Note that $E_j(\Phi)$ can be obtained from (5.6) while using (5.7) in (5.5). Then, the Huber M -estimator in (4.2) can be rewritten as

$$\min_{\Phi} \left\{ \mathcal{L} \equiv \sum_{\mathbf{x}} \left(\sum_{j=1}^J \rho(E_j(\Phi); \alpha_j) \right) \right\}. \quad (5.10)$$

Following [121], the Huber cost function, $\rho(\cdot)$, in (5.10) can be cast in a convex quadratic function as

$$\min_{\Phi} \left\{ \mathcal{L} = \frac{1}{2} \sum_{\mathbf{x}} \left(\sum_{j=1}^J F_j \cdot E_j^2(\Phi) + 2\alpha_j S_j \cdot E_j(\Phi) - \alpha_j^2 Q_j \right) \right\}. \quad (5.11)$$

The high-residual selective matrix is

$$S_j = \begin{bmatrix} s_{11} & \dots & s_{1M} \\ \vdots & \ddots & \vdots \\ s_{N1} & \dots & s_{NM} \end{bmatrix}, \quad s_{xy} = \begin{cases} -1, & E_j(\Phi; x, y) < -\alpha_j, \forall (x, y) \in R_j \\ 0, & |E_j(\Phi; x, y)| \leq \alpha_j, \forall (x, y) \in R_j \\ 1, & E_j(\Phi; x, y) > \alpha_j, \forall (x, y) \in R_j \\ d.c., & \forall (x, y) \notin R_j \end{cases}, \quad (5.12)$$

where *d.c.* denotes a “don’t care” value. The low-residual selective matrix is

$$F_j = \mathbf{1} - (S_j \cdot S_j), \quad \{\mathbf{1}, S_j, F_j\} \in \mathbb{R}^{N, M}. \quad (5.13)$$

We can estimate Φ using the Gauss-Newton algorithm [101] to solve the non-linear minimization problem in (5.11). Recall that $\hat{\Phi}$ is iteratively updated by (3.17) at each iteration g . Substituting (5.7) in (5.5), the residuals of each region, $E_j(\Phi)$, in (5.11) can be obtained from (5.6). As well, $E(\Phi)$ in (5.6) can be still approximated by (3.14) assuming affine motion model in (3.15). Replacing $E_j(\cdot)$ in (5.11) by its 1st order Taylor series expansion, then \mathcal{L} in (5.11) can be rewritten as

$$\mathcal{L} \approx \frac{1}{2} \sum_{\mathbf{x}} \sum_{j=1}^J \left[F_j \cdot (E_j(\Phi_{g-1}) + \Delta_g^\top E_j'(\Phi_{g-1}))^2 + 2\alpha_j S_j \cdot (E_j(\Phi_{g-1}) + \Delta_g^\top E_j'(\Phi_{g-1})) - \alpha_j^2 Q_j \right], \quad (5.14)$$

where $E'_j = \partial E_j / \partial \Phi$. Setting the gradient of \mathcal{L} w.r.t. Δ to zero, we obtain

$$-\sum_{j=1}^J [F_j \cdot (E_j(\Phi_{g-1}) E'_j(\Phi_{g-1})) + \alpha_j S_j \cdot E'_j(\Phi_{g-1})] = \hat{\Delta}_g^\top \sum_{j=1}^J [F_j \cdot (E'_j(\Phi_{g-1}) E_j^\top(\Phi_{g-1}))]. \quad (5.15)$$

We can write (5.15) in matrix notation as

$$-\mathbf{Y} \mathbf{P}^\top = (\mathbf{Y} \mathbf{H}^\top) \hat{\Delta}, \quad (5.16)$$

where

$$\mathbf{H} = [H_{1,1}, H_{1,2}, \dots, H_{N,M}], \quad (5.17)$$

$$\mathbf{Y} = [Y_{1,1}, Y_{1,2}, \dots, Y_{N,M}], \quad (5.18)$$

$$\mathbf{P} = [P_{1,1}, P_{1,2}, \dots, P_{N,M}], \quad (5.19)$$

$$H_{n,m} = \sum_{j=1}^J Q_j \cdot [nI_x, mI_x, nI_y, mI_y, I_x, I_y, -I_1 Q_1, \dots, -I_1 Q_J, -Q_1, \dots, -Q_J]^\top, \quad (5.20)$$

$$Y_{n,m} = f_{nm} \sum_{j=1}^J Q_j \cdot [nI_x, mI_x, nI_y, mI_y, I_x, I_y, -I_1 Q_1, \dots, -I_1 Q_J, -Q_1, \dots, -Q_J]^\top, \quad (5.21)$$

$$\text{and } \mathbf{P} = \text{ColVector} \left[\sum_{j=1}^J (\alpha_j S_j + F_j \cdot I_2) \right], \quad (5.22)$$

where $I_x = \partial I_2 / \partial x$, $I_y = \partial I_2 / \partial y$, “ColVector” refers to reshaping the resulting matrix into a column vector, and f_{nm} is an element located at position (n, m) of matrix F that can be stated as

$$F = \sum_{j=1}^J F_j. \quad (5.23)$$

It can be shown that equations (5.16) through (5.23) can be used to perform one iteration for finding a solution of $\hat{\Delta}$. The unknown vector, Φ , is iteratively accumulated due to (3.17) resulting in a single final estimate. The approach presented in this section can be referred to as weighted region-based Huber M -estimation for arbitrarily-shaped local illumination variations, WH-ASLIV_{6,2J}. The next section discusses the implementation setup performed to develop the experiments.

5.4 Implementation Setup

First, we would like to state that we compared the proposed approach, WH-ASLIV, to:

1. an approach that employs the global illumination model with a least-squares estimator as presented in [10, 11], and its implementation can be found online at [107]; referred to as LS-GIM,
2. a second approach that employs the affine illumination model with a least-squares estimator as presented in [13]; referred to as LS-AIM,
3. a third approach that exploits the ASLIV registration model as presented in Chapter 3 and in [18, 19] with a least-squares estimator; referred to as LS-ASLIV,
4. a fourth approach that uses the ASLIV model with the ordinary formula of Huber M -estimator as introduced in Chapter 4 and in [20]; referred to as HM-ASLIV, and
5. a fifth approach that is identical to WH-ASLIV approach, except that all weights in W_j are set to 1 (*i.e.*, Section 5.2.1 only). This approach can be referred to as RH-ASLIV, region-based Huber.

Note that all ASLIV-based approaches are suffixed by ‘6,2 J ’ according to J . It is worth noting that the LS-ASLIV_{6,2} corresponds to the LS-GIM in [10, 11], however, we use the code in [107].

The six implementations run on a 2 GHz Pentium IV Core 2 Duo, with 2 GB of RAM. All six approaches iteratively exploit a coarse-to-fine framework with 5 resolution levels (*i.e.*, $r=5$). Consequently, the unknown vector $\hat{\Phi}$ is adequately estimated at the coarsest level for both real and simulated data sets. For the natural canyon pair, the resolution level is set to 3 (*i.e.*, $r=3$) so as to still retain adequate estimates. Also, the implementation parameters are set as mentioned in Section 3.7, *i.e.*, the stopping threshold $\varepsilon=0.1$, the maximum number of iterations $g=10$. As well, Φ° is initialized either by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0]$ using the ASLIV_{6,6}-based approaches or by $[1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0]$ with the ASLIV_{6,8}-based approaches.

There are many ways to choose the threshold, α_j , for the Huber function in (5.11) as previously shown in Section 4.5. We still set α_j to $1.345\sigma_j$ [17, 101], where σ_j is the standard deviation of the residuals of each illumination region R_j , eliminating 5% of the outliers assuming Gaussian distribution on the residuals. Note that α_j is employed in both the RH-ASLIV and WH-ASLIV approaches. Similarly, the threshold α , used in the HM-ASLIV approach [20], can be set to 1.345σ as shown in Section 4.5.

It’s worth noting that the proposed WH-ASLIV approach assumes that the boundary layer of each illumination region consists of T pixels. Since, in a real imaging scenario, these regions have different sizes, the thickness T can’t hold for all of them. Therefore, the WH-ASLIV approach employs the Tukey function with $T=8$ for the simulated data sets and $T \in \{2, 3, 4\}$ for the real data set (these values of T are empirically chosen to obtain the most improved GRP).

While using a segmentation algorithm in the proposed approach, WH-ASLIV, some deviations would occur in each illumination region boundary. In real imaging scenarios, these deviations can't be precisely determined. Thus, we need to run an experiment to test the robustness of the WH-ASLIV approach against segmentation errors as opposed to competing approaches using simulated data sets. Some morphological operations (*i.e.*, dilation and erosion) can be applied to the boundary of each region to simulate such segmentation deviations. Each deviation, referred to segmentation noise level (SNL), can be simulated as follows. First, a dilation step is applied along the mask boundary. Some of the dilated pixels around the region boundary are randomly kept. Then, in a similar manner, an erosion step is performed on the region boundary. As well, some of the eroded pixels around the region boundary are randomly removed. Those two steps, dilation and erosion, are repeated for a number of iterations. We define each SNL by the summation of two parameters: P_d and P_e that denote the number of dilated and eroded pixels, respectively. The results of this experiment will be shown in Section 5.6.3.

5.5 Summary on The Proposed WH-ASLIV Approach

Thus far, we presented the weighted region-based registration approach, WH-ASLIV, using distinct M -estimators; each has its own cost function (*i.e.*, the Huber function). As well, the proposed approach, WH-ASLIV, imposes weights using a weighting function (*i.e.*, Tukey function) to lessen the impacts of the mis-classified areas located on the boundary of the illumination regions. Algorithm 4.1 summarizes the proposed approach WH-ASLIV that is cast in an iterative coarse-to-fine framework.

Algorithm 5.1 The proposed image registration approach WH-ASLIV

-
- 1: **Given:** two input images, I_1 and I_2 , as described in Section 3.3.
 - 2: **Required:** vector, Φ , that includes the finally estimated geometric and illumination parameters.
 - 3: **Initialize:** $\Phi \dots$ the unknown vector by Φ° ,
 - 4: $r \dots$ the no. of resolution levels,
 - 5: $g \dots$ the max no. iterations per resolution level,
 - 6: $\varepsilon \dots$ a predefined threshold of the cost function update,
 - 7: $J \dots$ the no. of illumination regions,
 - 8: $i \dots$ iteration counter by 1, and
 - 9: $T \dots$ the thickness of a boundary layer of an illumination region.
 - 10: **for** $p = 1$ to r **do**
 - 11: **repeat**
 - 12: Apply $\Phi^{(i)}$ to I_1 .
 - 13: Determine the AID⁽ⁱ⁾: $I_{21} = |I_2 - I_1^{\Phi^{(i)}}|$.
 - 14: Apply the k -means algorithm to the AID⁽ⁱ⁾ to obtain the binary masks, $Q_j^{(i)}$, of the illumination regions setting k to J .
 - 15: Determine the residuals, $E(\Phi^{(i)})$, using (3.14).
 - 16: Determine the boundary layer, D_j , of thickness T pixels for each illumination region.
 - 17: Determine the weighting matrix, W_j , of each illumination region that assign weights to its boundary layer using (5.5) with a weighting function (*i.e.*, Tukey function).
 - 18: Determine the residual, $E_j^{(i)}$, of each illumination region using (5.6).
 - 19: Determine the standard deviation, $\sigma_j^{(i)}$, of the residuals, $E_j^{(i)}$, of each illumination region.
 - 20: Set the cost function's (*i.e.*, the Huber function) threshold, $\alpha_j^{(i)}$, of each illumination region to $1.345 \sigma_j^{(i)}$ [17, 101].
 - 21: Determine the large-residual selective matrix, $S_j^{(i)}$, of each illumination region using (5.12).
 - 22: Determine the small-residual selective matrix, $F_j^{(i)}$, of each illumination region using (5.13).
 - 23: Determine the corresponding value of the cost function, $\mathcal{L}(\Phi^{(i)})$, using (5.14).
 - 24: Determine the cost function update $\delta_i = \mathcal{L}(\Phi^{(i)}) - \mathcal{L}(\Phi^{(i-1)})$.
 - 25: Determine the unknown vector's update, $\Delta^{(i)}$, using (5.16).
 - 26: Update $\Phi^{(i)}$ by $\Delta^{(i)}$ using (3.17).
 - 27: $i \leftarrow i + 1$.
 - 28: **until** ($\delta^{(i)} < \varepsilon$) or ($i > g$)
 - 29: **end for**
-

5.6 Experiments & Results

In this section, we report the experimental results performed using competing approaches applied on both simulated and real data in Section 5.6.1 and Section 5.6.2, respectively. As well, Section 5.6.3 presents an experiment showing the resistance of the proposed approach against simulated segmentation deviations.

5.6.1 Experiments on Simulated Data Sets

The first set of experiments is performed on the simulated data sets “ $J=3$ ” and “ $J=4$ ” data set using the LS-GIM, LS-AIM, LS-ASLIV, HM-ASLIV, RH-ASLIV and WH-ASLIV approaches considering the ASLIV version.

Table 5.1 show that the proposed approach, WH-ASLIV_{6,6}, provides more precise motion estimates compared to the LS-GIM, LS-AIM, LS-ASLIV_{6,6}, HM-ASLIV_{6,6} and RH-ASLIV_{6,6} approaches by an average of 90.4%, 86.5%, 33.7%, 18.4% and 5.9%, respectively, using the “ $J=3$ ” data set. Similarly, with “ $J=4$ ” data set, the proposed approach, WH-ASLIV_{6,8}, provides more precise motion estimates compared to the LS-GIM, LS-AIM, LS-ASLIV_{6,8}, HM-ASLIV_{6,8} and RH-ASLIV_{6,8} approaches by an average of 86.9%, 81.5%, 29.2%, 13.8% and 5.6%, respectively, as shown in Table 5.2. Generally speaking, the proposed approach yields better improved GRP compared to competing approaches.

To show the efficiency of WH-ASLIV approach, let us have a look at Table 5.3. This table reports SSIM indexes using the LS-GIM, LS-ASLIV_{6,6} and HM-ASLIV_{6,6}, RH-ASLIV_{6,6} and WH-ASLIV_{6,6} approaches with “ $J=3$ ” data set. The WH-ASLIV_{6,6} approach outperforms the other approaches by an average of 14.8%, 11.2% 2.6%, 1.2% and 0.7%, respectively.

As well, Table 5.4 shows that the proposed approach outperforms the competing approaches by an average of 27.5%, 23.4%, 5.3%, 3.2% and 1.2%, respectively, in terms of PSNR. Recall that these image quality measures express the correlation between the overlapping region of the registered pairs. As well, one can notice a reasonable increase in computational time using WH-ASLIV approach by an average of 33.1%, 29.2%, 11.6%, 4.7% and 0.5% compared to other approaches, respectively, as shown in Table 5.5.

Similarly, Table 5.6 shows the SSIM indexes using the LS-GIM, LS-AIM, LS-ASLIV_{6,8} and HM-ASLIV_{6,8}, RH-ASLIV_{6,8} and WH-ASLIV_{6,8} approaches with “ $J=4$ ” data set. The WH-ASLIV_{6,8} approach outperforms the other approaches by an average of 14.5%, 11.5%, 2.4%, 1.1% and 0.6%, respectively.

As well, Table 5.7 shows that the proposed approach outperforms the competing approaches by an average of 27.1%, 23.3%, 5.1%, 2.9% and 1.0%, respectively, in terms of PSNR. Recall that these image quality measures express the correlation between the overlapping region of the registered pairs. In addition, a reasonable increase in computational time can be noticed with the WH-ASLIV approach by an average of 33.4%, 30.5% 12.0%, 5.1% and 0.9%, respectively, compared to competing approaches, as shown in Table 5.8.

Fig. 5.3(a-f) shows the normalized AIDs of the aligned pairs of the “ $J=3$ ” simulated pair shown in Fig. 3.4(a,b) using competing approaches, yielding SSIM = 0.9642, 0.9691, 0.9789, 0.9802, 0.9818 and 0.9837, and PSNR = 21.91, 23.35, 27.01, 27.76, 28.45 and 28.87 dB, respectively. It is worth noting that many areas in Fig. 5.3(f) are darker than their correspondences in Fig. 5.3(a-e), especially inside each illumination region and around its boundary layer.

Table 5.1: The AAEs ($\times 10^{-4}$) of the estimated affine parameters of the “ $J=3$ ” data set. The lower the value of AAEs, the better the approach performance.

	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}	RH-ASLIV _{6,6}	WH-ASLIV _{6,6}
a_1	15.3658	12.1419	1.7205	1.2025	0.9806	0.9167
a_2	12.0017	10.5204	1.5079	1.0344	0.8096	0.7748
a_3	14.6081	11.9571	1.6189	1.1661	0.9399	0.8779
a_4	19.1128	15.8924	2.0502	1.8699	1.6531	1.5612
a_5	9190.76	7649.53	789.16	722.35	691.29	647.22
a_6	2688.72	2236.18	795.70	716.11	685.77	644.32

Table 5.2: The AAEs ($\times 10^{-4}$) of the estimated affine parameters of the “ $J=4$ ” data set. The lower the value of AAEs, the better the approach performance.

	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}	RH-ASLIV _{6,8}	WH-ASLIV _{6,8}
a_1	16.5962	13.7167	1.8121	1.4358	1.2694	1.2046
a_2	13.2015	10.6395	1.6025	1.1916	1.0178	0.9684
a_3	15.5246	13.2940	1.7360	1.2049	1.0653	0.9937
a_4	21.0259	17.3674	2.4261	1.9280	1.7458	1.6283
a_5	9472.09	7918.66	816.50	760.57	742.91	702.39
a_6	2835.12	2382.50	804.81	755.11	739.20	699.93

Table 5.3: SSIM for a random subset of the “ $J=3$ ” data set.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}	RH-ASLIV _{6,6}	WH-ASLIV _{6,6}
10	0.8397	0.8530	0.9602	0.9622	0.9690	0.9745
17	0.8194	0.8379	0.9577	0.9589	0.9630	0.9698
28	0.8268	0.8645	0.9534	0.9598	0.9664	0.9733
34	0.8066	0.8358	0.9507	0.9531	0.9590	0.9667
38	0.8552	0.8853	0.9650	0.9665	0.9709	0.9781
42	0.8201	0.8413	0.9612	0.9681	0.9736	0.9802
48	0.8419	0.8742	0.9634	0.9653	0.9695	0.9751

Table 5.4: PSNR for a random subset of the “ $J=3$ ” data set.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}	RH-ASLIV _{6,6}	WH-ASLIV _{6,6}
10	22.53	24.79	29.02	29.78	30.21	30.53
17	20.01	22.14	28.33	29.01	29.48	29.87
28	21.76	24.12	29.42	30.17	30.77	31.12
34	21.91	23.35	27.01	27.76	28.45	28.87
38	23.47	24.93	28.63	29.25	29.82	30.36
42	22.95	24.89	29.25	29.87	30.48	30.71
48	21.94	22.05	29.69	30.10	30.55	30.90

Table 5.5: Computational time (sec.) for a random subset of the “ $J=3$ ” data set.

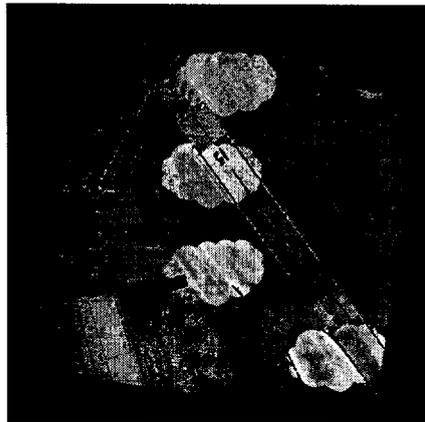
#	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}	RH-ASLIV _{6,6}	WH-ASLIV _{6,6}
10	43.0	44.6	57.7	62.2	64.3	64.6
17	42.8	43.8	55.4	59.8	62.4	62.8
28	44.7	46.0	58.8	61.4	63.9	64.3
34	43.3	44.9	57.2	62.5	65.3	65.6
38	42.7	44.1	56.4	63.2	66.1	66.4
42	43.4	45.2	56.5	60.0	62.8	63.2
48	41.2	43.4	55.3	59.4	62.4	62.7

5.6.2 Experiments on the Real Data Set

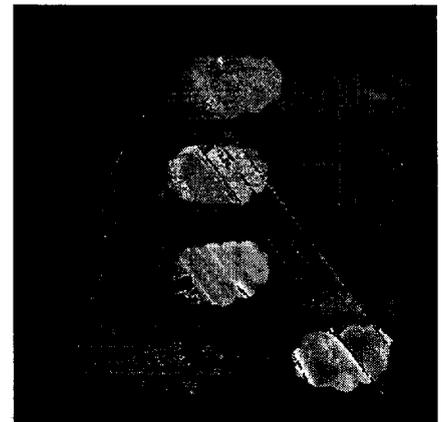
Our next set of experiments were performed on the real data set as described in Section 3.5 using LS-GIM, LS-AIM, LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6} and WH-ASLIV_{6,6} approaches with the k -means algorithm for segmentation setting, with $k=J=3$.

Table 5.6: SSIM for a random subset of the “ $J=4$ ” data set.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}	RH-ASLIV _{6,8}	WH-ASLIV _{6,8}
10	0.7969	0.8295	0.9341	0.9417	0.9466	0.9513
17	0.7844	0.8174	0.9254	0.9363	0.9425	0.9566
28	0.8021	0.8348	0.9164	0.9227	0.9291	0.9405
34	0.7721	0.8039	0.9142	0.9253	0.9325	0.9339
38	0.8441	0.8762	0.9284	0.9494	0.9501	0.9559
42	0.7778	0.8103	0.9334	0.9405	0.9478	0.9672
48	0.7999	0.8329	0.9245	0.9306	0.9478	0.9505



(a) LS-GIM; PSNR=21.91 dB



(b) LS-AIM; PSNR=23.35 dB

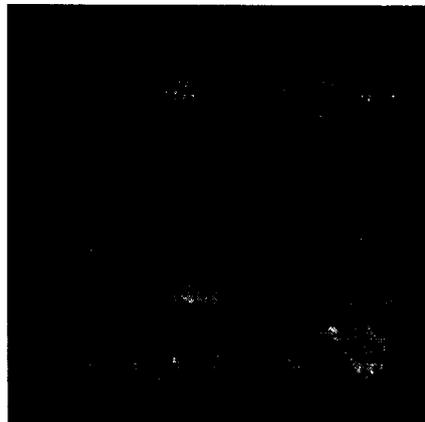
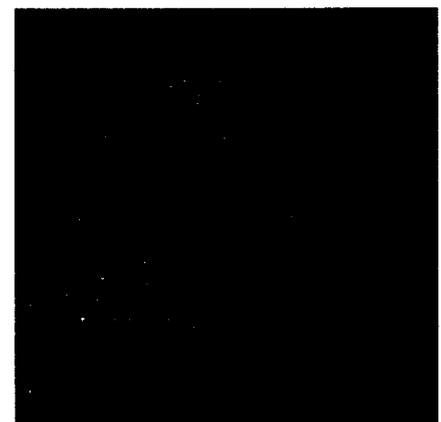
(c) LS-ASLIV_{6,6}; PSNR=27.01 dB(d) HM-ASLIV_{6,6}; PSNR=27.76 dB(e) RH-ASLIV_{6,6}; PSNR=28.45 dB(f) WH-ASLIV_{6,6}; PSNR=28.87 dB

Figure 5.3: (a-f) Normalized AIDs of the resulting aligned pairs for the “ $J=3$ ” simulated pair in Fig. 3.4(a,b); pair #34 in Table ???. The darker the AID, the better the approach performance.

Table 5.7: PSNR for a random subset of the “ $J=4$ ” data set.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}	RH-ASLIV _{6,8}	WH-ASLIV _{6,8}
10	22.12	23.97	28.57	29.15	29.69	30.13
17	19.71	21.75	27.66	28.35	29.01	29.50
28	21.37	23.42	28.95	29.76	30.15	30.79
34	21.51	23.36	26.66	27.46	27.79	28.50
38	22.99	24.83	27.93	28.77	29.24	30.06
42	22.60	24.75	28.62	29.39	29.97	30.29
48	21.67	23.81	29.18	29.82	30.15	30.32

Table 5.8: Computational time (sec.) for a random subset of the “ $J=4$ ” data set.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,8}	HM-ASLIV _{6,8}	RH-ASLIV _{6,8}	WH-ASLIV _{6,8}
10	43.4	45.8	59.0	63.7	65.6	66.5
17	44.0	47.3	57.3	60.9	63.7	64.5
28	46.6	49.3	60.2	62.8	65.2	65.7
34	45.4	48.1	58.8	63.8	66.9	67.1
38	44.5	47.8	58.0	65.1	67.2	68.0
42	43.9	46.5	57.7	61.1	64.2	64.8
48	42.9	45.4	56.8	60.7	64.2	63.8

Table 5.9 shows NCC values that express the correlation between the overlapping areas of the two aligned images using competing approaches. It can be shown that the WH-ASLIV approach surpasses other competing approaches. Note that the boundary layer thickness, T , is set to one value out of $\{2,3,4\}$ for each real image pair to provide the best results using WH-ASLIV approach. These T values are empirically obtained to achieve higher results. We can not constrain the parameter T to a certain value for all real image pairs as the illumination regions are not *a priori* known. However, these regions should be iteratively determined and refined using k -means segmentation algorithm.

Fig. 5.4(a-f) shows the gamma-corrected AIDs of the registered pairs, of the real pair in Fig. 2.1(c,d) using the LS-GIM, LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6} and WH-ASLIV_{6,6} approaches, with NCC = 0.9642, 0.9691, 0.9789, 0.9802, 0.9818

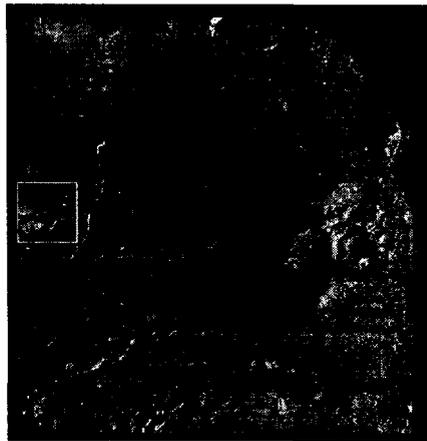
Table 5.9: NCC values for real data set, $k=J=3$. Note that pair #A and #B refer to Fig. 2.1(a,b) and (c,d), respectively (T denotes the boundary layer thickness). Recall that the higher the NCC value, the better the performance of the approach.

#	LS-GIM	LS-AIM	LS-ASLIV _{6,6}	HM-ASLIV _{6,6}	RH-ASLIV _{6,6}	WH-ASLIV _{6,6}
A	0.8819	0.8998	0.9226	0.9237	0.9261	0.9280 , $T=3$
B	0.9642	0.9691	0.9789	0.9802	0.9818	0.9837 , $T=3$
C	0.9799	0.9813	0.9853	0.9870	0.9891	0.9909 , $T=4$
D	0.9450	0.9520	0.9644	0.9659	0.9681	0.9704 , $T=2$
E	0.9601	0.9665	0.9743	0.9758	0.9782	0.9806 , $T=3$
F	0.9719	0.9786	0.9922	0.9939	0.9951	0.9962 , $T=2$
G	0.9551	0.9594	0.9657	0.9671	0.9690	0.9711 , $T=4$
H	0.8971	0.9087	0.9229	0.9245	0.9266	0.9283 , $T=3$
I	0.9081	0.9105	0.9241	0.9286	0.9305	0.9327 , $T=2$
J	0.9542	0.9604	0.9770	0.9833	0.9854	0.9878 , $T=4$

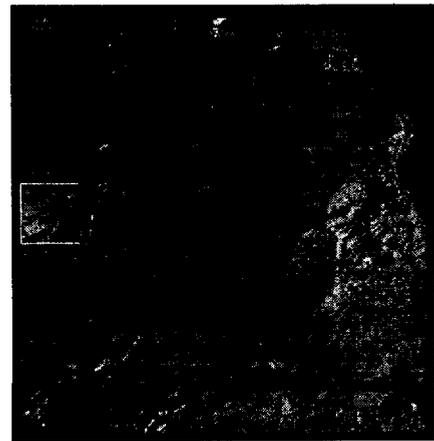
and 0.9837, respectively ($\gamma=1.7$ for better visualization, $k=J=3$). One can notice that Fig. 5.4(f) has many areas that are darker than the corresponding areas in Fig. 5.4(a-e). Enlarged views of the white boxes in Fig. 5.4(a-f) are shown in Fig. 5.5(a-f), respectively. This real example gives an impression on how well the WH-ASLIV approach yields improvements in GRP and illumination correction.

5.6.3 Experiments on Simulated Segmentation Deviations

In this subsection, we demonstrate the robustness of the WH-ASLIV approach against possible segmentation deviations as opposed to competing ones. Changing the segmentation deviations in real imaging is unavailable as there is no *a priori* information on the exact illumination region segmentation. Therefore, some morphological operations (dilation and erosion) are applied on the boundary of



(a) LS-GIM; NCC=0.9642



(b) LS-AIM; NCC=0.9691

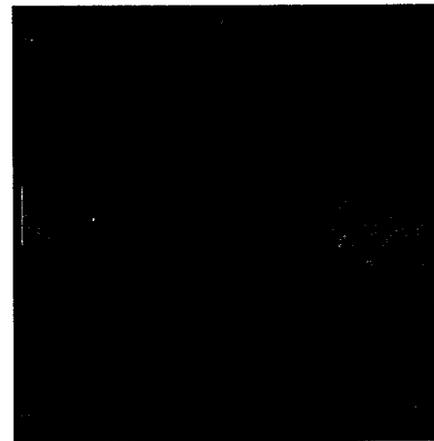
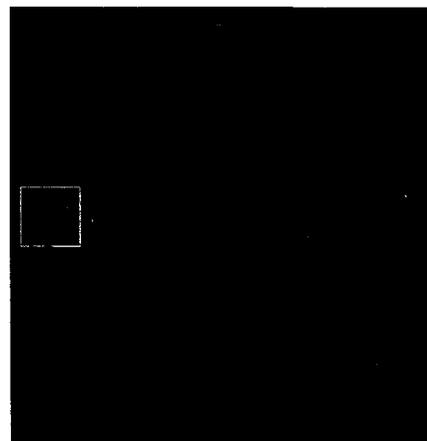
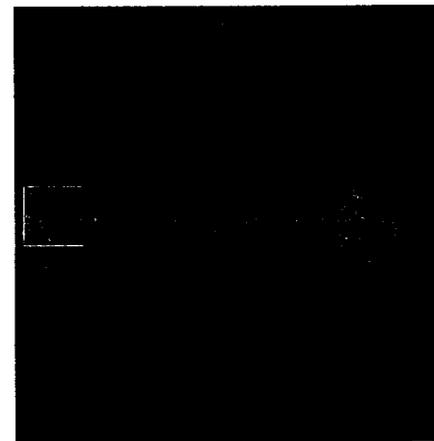
(c) LS-ASLIV_{6,6}; NCC=0.9789(d) HM-ASLIV_{6,6}; NCC=0.9802(e) RH-ASLIV_{6,6}; NCC=0.9818(f) WH-ASLIV_{6,6}; NCC=0.9837

Figure 5.4: (The gamma-corrected AIDs for the real pair in Fig. 2.1(c,d), $\gamma = 1.7$. The darker the AID, the better the approach performance.



(a) LS-GIM



(b) LS-AIM

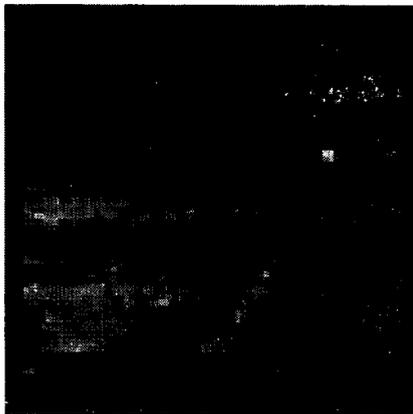
(c) LS-ASLIV_{6,6}(d) HM-ASLIV_{6,6}(e) RH-ASLIV_{6,6}(f) WH-ASLIV_{6,6}

Figure 5.5: (a-f) Enlarged views of the white boxes in Fig. 5.4(a-f), respectively, $\gamma = 1.7$ and $J=3$. The darker the AID, the better the approach performance.

the *a priori* illumination regions in the simulated data sets to simulate the segmentation deviations (see Section 5.4).

This experiment is performed on the “ $J=3$ ” data set using multiple segmentation noise levels (SNLs) using the LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6}, and WH-ASLIV_{6,6} approaches. Fig. 5.6 shows the absolute errors (AEs) of the estimated affine parameters compared to their ground truth values of the six affine parameters versus multiple SNLs. Note that each SNL corresponds to a certain summation of eroded and dilated pixels ($P_e + P_d$). The LS-GIM is considered in the experiment, however, it is not shown in the corresponding figure for better visualization. Throughout Fig. 5.6, one can notice that the AE using the WH-ASLIV approach is significantly less than those of using other approaches, while increasing the SNL. One can also observe that the slope of the AE using WH-ASLIV approach is less than those of other approaches while the SNL increases. This observation means that the proposed approach, WH-ASLIV, provides more resistance than competing approaches versus segmentation deviations by an average increase of 90.8%, 83.4%, 36.5%, 21.0% and 7.6%, respectively.

Furthermore, Tables (B.1) through (B.6), shown in Appendix B, report the percentage increase of the AAEs of the estimated six affine parameters, respectively, for different SNLs w.r.t. those AAEs without segmentation error, *i.e.*, at SNL = 0, using LS-GIM, LS-AIM, LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6} and WH-ASLIV_{6,6} approaches for the “ $J=3$ ” data set. Depicted is that the WH-ASLIV approach outperforms other approaches while SNL increases, thereby being less sensitive to segmentation perturbations than the others. The results of those six tables are depicted in Fig. 5.6, respectively.

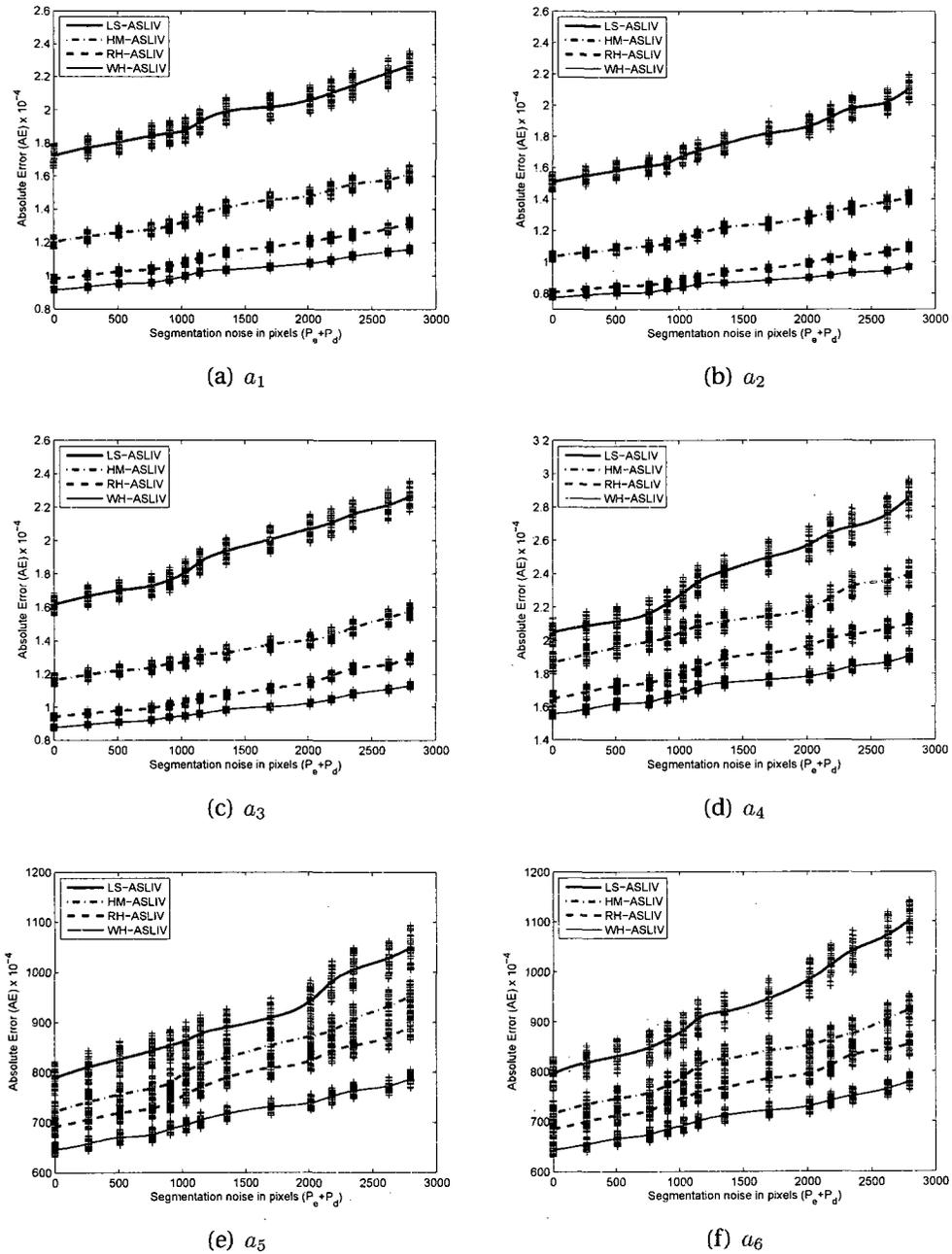


Figure 5.6: (a-f) The absolute error (AE) of six affine parameters using the LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6}, and WH-ASLIV_{6,6} approaches, respectively, at different segmentation noise levels. Both LS-GIM and LS-AIM approaches are not included for better visualization. The lower the value of AEs, the better the approach performance.

5.7 Analysis & Discussion

In this chapter, we present an image registration approach to cope with images having arbitrarily-shaped distinct illumination variations. In general, the proposed approach exploits different M -estimators. In theory, each M -estimator employs a distinct cost function with certain properties. It is preferable for a candidate cost function to have a threshold or a tuning parameter by which the residual's penalty could be controlled. If an objective function is to handle an illumination region, the threshold of that function would segment this region into small and large residuals. For example, the Huber function is selected as a cost function. Then, we extend its ordinary formula from one-threshold to multi-threshold, thus allowing to assign a proper tuning parameter to each illumination region. Thereby, the M -estimator linearly and quadratically penalizes the resulting large and small residuals. It can be shown that the proposed approach overcomes: i) the sensitivity of LSE by using a more robust M -estimator, and ii) the issue of iteratively applying one threshold over the whole residuals' space.

The proposed approach is cast in an iterative coarse-to-fine scheme. Mis-segmented pixels located on the boundary of illumination regions could result in two cases: i) if the currently estimates are applied at certain iteration, or ii) the binary masks which correspond to illumination regions are upscaled to the finer resolution level. Therefore, the proposed approach imposes a weighting function to decrease the impact of such pixels. The selected weighting function should be bounded with a saturation in order to still have full consideration to pixels that are correctly segmented. With no need for any feature-based approach for initialization, the proposed approach converges well using an identity initial guess. The GRP is significantly improved jointly with illumination correction using the

proposed approach as opposed to competing ones, on both real and simulated data sets with only reasonable increase in computational time.

Despite of the robustness of the proposed approach against arbitrary-shaped illumination variations and segmentation deviations, the approach is limited by certain constraints, such as

- A pin-hole camera model is assumed.
- Using a perfect camera is assumed, so that photometric corruptions are not considered.
- The input images are captured following the conditions mentioned in Section 3.2.
- The scene is assumed to have no object motion (*i.e.*, a static scene, but dynamic in terms of various illuminations).
- The proposed model, ASLIV, has piece-wise gain and offset.
- The distance between the camera and the scene as well as the light source(s) is assumed to be large so that the camera's focal length effect can be ignored.

The proposed approach is formulated such that many functions could be selected for many of the subparts. However, to implement the resulting mathematical equations, some functions and models have been specified, such as

- Affine motion model: that is extensively used by competing approaches and simple to implement.
- Huber function: that has an iteratively adaptive threshold by which the residuals could be controlled and differently penalized in cooperation with M -estimators.

- Tukey function: that has a tuning parameter by which the boundary layer thickness could be directly controlled so as to reduce the negative effects of the possibly mis-segmented pixels.

In the next chapter, we not only conclude our thesis research, but also clearly state the contributions, our own publications, and future work directions.

Chapter 6

Conclusions & Contributions

In this chapter, we present the research conclusions in Section 6.1. Then, the main contributions of this dissertation are clearly given in Section 6.2. The publications resulting from this research are mentioned in Section 6.3. Finally, possible future work directions are provided in Section 6.4.

Before we present the conclusions of this dissertation, we strongly emphasize that the principle target of this research is to “*improve the sub-pixel geometric registration precision*” by minimizing the factors affecting this process, such as illumination variations. Therefore, we give a summary on the work presented in this dissertation and comment to the results of the experiments that are also developed.

Experiments are performed on real image pairs as well as simulated data sets to allow ground truth verification of the estimated parameters. Normalized cross-correlation (NCC), structural similarity (SSIM) index, and peak signal-to-noise ratio (PSNR) are used to evaluate the performance of the competing approaches. The proposed approach, WH-ASLIV, is compared to: (i) the global illumination approach [10, 11], (ii) the affine illumination approach [13], (iii) the LS-ASLIV approach [18, 19], (iv) the approach using the Huber M -estimator with the

ASLIV model [20] and (v) the approach using region-based Huber M -estimators with the ASLIV model.

- In terms of sub-pixel geometric registration precision, clear improvements are obtained using the proposed approach, WH-ASLIV, compared to competing approaches by an average increase of 90.1%, 84.6%, 31.5%, 16.1% and 5.8%, respectively, with a reasonable average increase of 42.1%, 36.3%, 13.0%, 4.7% and 0.5%, respectively, in computational time.
- In terms of the selected image quality metrics, the proposed approach, WH-ASLIV, outperforms competing approaches by an average increase of 17.4%, 14.2%, 3.4%, 1.3% and 0.7% in SSIM, respectively, and average increase of 36.9%, 32.5%, 5.7%, 3.1% and 1.2% in PSNR, respectively.
- The proposed approach, WH-ASLIV, provides more robustness against segmentation errors by an average increase of 90.8%, 83.4%, 36.5%, 21.0% and 7.6%, respectively, compared to competing approaches.

6.1 Conclusions

In this section, we can summarize the conclusions as follows:

- A registration model that copes with images having arbitrarily-shaped locally variant illuminations, results in an improvement in geometric registration precision (GRP) and illumination correction.
- Choosing a robust estimator improves the GRP.
- If the candidate cost function has certain properties, especially a tuning parameter, this leads to a more precise GRP.

- The impacts of the mis-segmented areas, located on the boundary of each illumination region due to using a clustering algorithm, should be reduced, yielding more accurate GRP.
- Employing a coarse-to-fine framework allows to cope with large deformations, obtain accurate estimates and save the computational time.

6.2 Contributions

In this section, we can summarize the main contributions of this dissertation:

1. Robustness against local illumination changes:

We proposed an intensity-based model to obtain a more accurate sub-pixel geometric registration jointly with illumination correction. This model is robust against locally variant illumination regions in the images and outperforms the global illumination model in terms of geometric registration precision as shown in [19]. A special case, when considering only two illumination regions, is published in [18].

2. Capability of differently penalizing local residuals:

The proposed approach simultaneously employs different robust M -estimators. We extend the ordinary formula of an objective function having single-threshold to be a spatially adaptive region-based function with multiple-threshold. Thus, each threshold is iteratively assigned to an illumination region. A framework of spatially adaptive M -estimators is achieved using a cost function with certain properties. A special case, when using the ordinary Huber function as cost function, is accepted for publication in [20].

3. Robustness against segmentation errors:

The proposed approach exhibits more resistance versus mis-segmented areas located on the boundary layer of each illumination region due to segmentation. The proposed approach is able to reduce the negative effects of those mis-segmented areas by lessening their contributions throughout the registration process using a weighting function. Thus, more precise geometric registration precision as well as illumination correction is obtained. This contribution and its precedent are submitted to [22].

6.3 Publications

In this section, we state the published, accepted and submitted manuscripts out of this research as follows:

1. M. M. Fouad, R. M. Dansereau, A. Whitehead, "Geometric registration of images with arbitrarily-shaped local intensity variations from shadows," in *Proc. of IEEE Intern. Conf. on Image Processing (ICIP'2009, Cairo, Egypt)*, November 7-11, 2009, pages 201-204.
2. M. M. Fouad, R. M. Dansereau, and A. D. Whitehead, "Geometric image registration under locally variant illuminations using Huber M-estimator," A. Elmoataz et al. (Eds.): *(ICISP'2010, Trois-Rivières, QC, Canada)*, LNCS 6134, pages 10-18. Springer, Heidelberg, 2010.
3. M. M. Fouad, R. M. Dansereau, and A. D. Whitehead, Geometric Image Registration Under Arbitrarily-Shaped Locally Variant Illuminations, *Signal, Image & Video Processing*, Springer, London, July 2010, accepted.

4. M. M. Fouad, R. M. Dansereau, and A. D. Whitehead, Image Registration Under Illumination Variations Using Weighted Region-based M -estimator, submitted to *IEEE Trans. on Image Processing*.

6.4 Future Work

Though the proposed approach works well with images that follows certain conditions (see Section 3.2) under limitations as mentioned in the last part of Section 5.7, the proposed approach does not succeed if, *e.g.*, indoor images are considered or one of those limitations is violated. Thus, there is still room for improvements.

The input images are assumed to be taken with the same exposure. In turn, the proposed approach will not work well if, *e.g.*, photometric corruptions are involved, such as sensor noise, lens distortion and atmospheric scintillations. This issue can be tackled by starting with embedding one photometric corruption simultaneously with both illumination variations and geometric deformations. The constraints would then be changed to produce more precise estimates. However, the combination of two non-geometric corruptions with geometric transformations must be done carefully. Otherwise, it would result in nonconvergence or a significant increase of the computation time.

As well, the light source is assumed to be far from the scene. Thus, the proposed model, ASLIV, is assumed to have piece-wise gain and offset. However, this model is not sufficient if near-field light source(s) exist w.r.t. the scene. In such a case, the brightness uniformity along with each illumination region would degrade the geometric registration precision. One remedy is to model the local illumination changes in a more generalized illumination model, such as affine

illumination model along with each distinct illumination region.

Besides, the camera is assumed to be far from the scene (*i.e.*, outdoor images result in). In turn, the dimensions of the objects in the scene could comparably neglected. Thus, a 6-parameter affine motion model would be reasonable to represent the geometric deformations between the captured images. However, the proposed approach will not even converge to a closest answer, if indoor images are considered. This issue arises because the focal length of the camera cannot be neglected in such an indoor images. One solution is to replace the 6-parameter affine motion model with a more generalized motion model, such as the 8-parameter perspective model. If this issue is carefully solved, an additional category of images can be considered.

Another important avenue of the work is to broaden the proposed approach for some other cost functions. We first use the quadratic cost function leading to a least-squares estimator. However, its derivative function gives an increasing weights to the residuals. Thus, the minimization problem becomes more sensitive to the outliers. Hence, the quadratic function has been replaced with a robust function, such as the Huber function. The advantage of the Huber function over the quadratic function is that the former assigns constant weight to the outliers as its first derivative becomes constant after a certain threshold. Thus, the sub-pixel geometric registration precision enhances. However, we should try whether the estimates get improved if the outliers are zero-weighted or not. That solution might be tackled if another cost function, such as Bisquare function has been used instead.

List of References

- [1] T. Q. Pham. *Spatiotonal Adaptivity in Super-Resolution of Under-sampled Image Sequences*. PhD thesis, Quantitative Imaging Group, University of Delft, Netherlands, October 2006.
- [2] B. Zitová and J. Flusser. Image registration methods: A Survey. *ELSEVIER J. on Image & Vision Computing*, 21:977–1000, June 2003.
- [3] L. Lucchese, S. Leorin, and G. M. Cortelazzo. Estimation of two-dimensional affine transformations through polar curve matching and its application to image mosaicking and remote-sensing data registration. *IEEE Trans. Image Processing*, 15(10):3008–3019, October 2006.
- [4] M. Gevrekci and B. K. Gunturk. Superresolution under photometric diversity of images. *Advances in Signal Processing*, April 2007.
- [5] Y. Wang and L. Xu. A global optimized registration algorithm for image stitching. In *IEEE CISP*, volume 3, pages 525–529, May 2008.
- [6] L. Lou, F. Zhang, C. Xu, F. Li, and M. Xue. Automatic registration of aerial image series using geometric invariance. In *IEEE ICAL*, pages 1198–1203, September 2005.
- [7] M. Beauchemin and K. B. Fung. An adaptive filter for the reduction of artifacts caused by image misregistration. In *Intr. Symp. Analysis of Multi-Temporal Remote Sensing Images*, pages 174–176, May 2005.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Medical Image Analysis*, 60(2):90–110, November 2004.
- [9] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1), December 2006.

-
- [10] S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Trans. Medical Imaging*, 22(7):865–874, July 2003.
 - [11] P. M. Q. Aguiar. Unsupervised simultaneous registration and exposure correction. In *IEEE ICIP*, pages 361–364, June 2006.
 - [12] A. Bartoli. Groupwise geometric and photometric direct image registration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(12):2098–2108, December 2008.
 - [13] Y. Altunbasak, R. M. Mersereau, and A. J. Patti. A fast parametric motion estimation algorithm with illumination and lens distortion correction. *IEEE Trans. Image Processing*, 12(4):395–408, April 2003.
 - [14] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: What is knowable? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(11):1455–1467, November 2003.
 - [15] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. Graphics*, 22(3):319–325, July 2003.
 - [16] P. J. Huber. *Robust Statistics*. Wiley, New York, U.S.A., first edition, 1981.
 - [17] P. J. Huber. *Robust Statistics Procedures*. SIAM, Bayreuth, Germany, second edition, 1996.
 - [18] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead. Geometric registration of images with arbitrarily-shaped local intensity variations from shadows. In *IEEE ICIP*, pages 201–204, November 2009.
 - [19] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead. Geometric image registration under arbitrarily-shaped locally variant illuminations. *Signal, Image & Video Processing*, July 2010. accepted.
 - [20] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead. Geometric image registration under locally variant illuminations using Huber M -estimator. In *Inter. Conf. on Image & Signal Processing*, volume LNCS 6134, pages 10–18. Springer, Heidelberg, June 2010.
 - [21] J. Jiang, S. Zheng, A. W. Toga, and Z. Tu. Learning based coarse-to-fine image registration. In *IEEE ICCVPR*, pages 1–7, June 2008.

-
- [22] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead. Image registration under illumination variations using weighted region-based m -estimators. *IEEE Trans. Image Processing*. submitted.
- [23] N. Álvarez, J. Sanchiz, J. Badenas, F. Pla, and G. Casañ. Contour-based image registration using mutual information. *Pattern Recognition and Image Analysis, Springer*, pages 227–234, May 2005.
- [24] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(3):347–364, March 2002.
- [25] S. Aylward, J. Jomier, S. Weeks, and E. Bullitt. Registration of vascular images. *Computer Vision*, 55(2):123–138, 2003.
- [26] D. Capel and A. Zisserman. Computer vision applied to super resolution. *IEEE Signal Processing Magazine*, 20(3):75–86, May 2003.
- [27] D. Capel. *Image Mosaicing and Super-resolution*. Distinguished dissertations. Springer, New York, 2004.
- [28] M. Gevrekci and B. K. Gunturk. On geometric and photometric registration of images. In *IEEE ICASSP*, volume 1, pages 1261–1264, April 2007.
- [29] P. Besl and N. McKay. A method for registration of 3D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [30] I. Zoghlami, O. Faugeras, and R. Deriche. Using geometric corners to build a 2D mosaic from a set of images. In *IEEE ICCVPR*, pages 420–425, June 1997.
- [31] J. K. Seo, G. C. Sharp, and S. W. Lee. Range data registration using photometric features. In *IEEE ICCVPR*, volume 2, pages 1140–1145, June 2005.
- [32] J. M. Fitzpatrick and J. B. West. The distribution of target registration error in rigid-body point-based registration. *IEEE Trans. Medical Imaging*, 20(9): 917–927, September 2001.
- [33] E. Guest, E. Berry, R. A. Baldock, M. Fidrich, and M. A. Smith. Robust point correspondence applied to 2D and 3D image registration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(2):165–179, February 2001.

-
- [34] Y. Dufournaud, C. Schmid, and R. Horaud. Image matching with scale adjustment. *Image and Vision Computing*, 93:175–194, October 2004.
- [35] H. Pottmann, S. Leopoldseder, and M. Hofer. Image matching with scale adjustment. *Computer Vision and Image Understanding*, 95(1):54–71, July 2004.
- [36] S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *Proc. of Seventh European Conference on Computer Vision*, volume 2353, pages 69–73, 2002.
- [37] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard. Unsupervised learning of an ATLAS from unlabeled point-sets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):160–172, February 2004.
- [38] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, New York, first edition, 1997.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, second edition, 2001.
- [40] H. Chui and A. Rangarajan. A feature registration framework using mixture models. In *Proc. Mathematical Methods in Biomedical Image Analysis*, volume 1, pages 190–197, June 2000.
- [41] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(3):114–141, 2003.
- [42] M. Brown and D. Lowe. Recognising panoramas. In *IEEE ICCV*, volume 2, pages 1218–1225, October 2003.
- [43] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Computer Vision*, 60(1):63–86, 2004.
- [44] M. Gevrekci and B. K. Guntürk. Reliable interest point detection under large illumination variations. In *IEEE ICIP*, pages 869–872, October 2008.
- [45] D. Xu and T. Kasparis. Robust image registration under spatially non-uniform brightness changes. In *IEEE ICASSP*, volume 2, pages 945–948, March 2005.

-
- [46] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *Proc. of Eighth European Conference on Computer Vision*, volume 2353, pages 69–73, 2004.
- [47] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, September 2004.
- [48] K. Mikolajczyk, T. Tuytelaars, C. Schmid, and A. Zisserman J. Matas. A comparison of affine region detectors. *Computer Vision*, 65(1):43–72, November 2005.
- [49] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.
- [50] L. V. Gool, T. Moons, and D. Ungureanu. Affine/photometric invariants for planar intensity patterns. In *Proc. of Fourth European Conference on Computer Vision*, volume 1064, pages 642–651, 1996.
- [51] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, June 2002.
- [52] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *IEEE CVPR*, volume 1, pages 510–517, June 2005.
- [53] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.
- [54] N. D. Cahill, J. A. Schnabel, J. A. Noble, and D. J. Hawkes. Revisiting overlap invariance in medical image alignment. In *IEEE ICCVPR*, pages 1–8, June 2008.
- [55] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Proc. of Computer Vision & Pattern Recognition*, volume 2, pages 506–513, 2004.
- [56] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato. Sift features tracking for video stabilization. In *Proc. 14th ICIAP*, pages 825–830, September 2007.

-
- [57] R. C. Bolles and M. A. Fischler. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *Proc. of Seventh Inter. Conference on Artificial Intelligence*, pages 637–643, May 1981.
- [58] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [59] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, U.S.A., second edition, 2003.
- [60] P. Perez and N. Garcia. Robust and accurate registration of images with unknown relative orientation and exposure. In *IEEE ICIP*, volume 3, pages 1104–1007, September 2005.
- [61] X. Llado, A. Del Bue, and L. Agapito. Recovering Euclidean deformable models from stereo-motion. In *IEEE ICPR*, pages 1–4, December 2008.
- [62] T. Zhao, M. Velliste, M. V. Boland, and R. F. Murphy. Object type recognition for automated analysis of protein subcellular location. *IEEE Trans. Image Processing*, 14(9):1351–1359, September 2005.
- [63] L. Ding, A. Goshtasby, and M. Satter. Volume image registration by template matching. *Image and Vision Computing*, 19:821–832, 2001.
- [64] H. Cheong, S. Park, M. Park, and K. Park. Vision-based global localization in indoor environment with an object entity-based hybrid map. In *IEEE ICCAS*, pages 218–223, October 2007.
- [65] J. Park, S. Kim, N. Doh, and K. Park. Vision-based global localization based on a hybrid map representation. In *IEEE ICCAS*, pages 1104–1108, October 2008.
- [66] L. Lucchese and G. M. Cortelazzo. A noise-robust frequency domain technique for estimating planar roto-translations. *IEEE Trans. Signal Processing*, 48(6):1769–1786, June 2000.
- [67] H. Foroosh, J. B. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Trans. Image Processing*, 11(3):188–200, March 2002.

-
- [68] M. Irani, E. Shterman, and Y. Caspi. Space-time super-resolution. *IEEE Trans. Pattern Analysis And Machine Intelligence*, 27(4):531–545, April 2005.
- [69] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Computer Vision*, 61(3): 211–231, 2005.
- [70] S. Ince and J. Konrad. Occlusion-aware optical flow estimation. *IEEE Trans. Image Processing*, 17(8):1443–1451, August 2008.
- [71] S. Zokai and G. Wolberg. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. *IEEE Trans. Image Processing*, 14(10):1422–1434, October 2005.
- [72] P. Vandewalle, S. Susstrunk, and M. Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *Applied Signal Processing*, pages 78–91, 2006.
- [73] W. S. Hoge. A subspace identification extension to the phase correlation method. *IEEE Trans. Medical Imaging*, 22(2):277–280, February 2003.
- [74] S. Kim and W. Su. Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Trans. Image Processing*, 2:534–539, October 1993.
- [75] P. Vandewalle, S. Susstrunk, and M. Vetterli. Double resolution from a set of aliased images. In *SPIE/IS&T Electronic Imaging*, volume 5301, pages 374–382, January 2004.
- [76] P. Vandewalle. *Super-resolution from unregistered aliased images*. PhD thesis, Section Systemes De Communication, EPFL, Laussane, Belgium, July 2006.
- [77] B. Lucas and T. Kanade. An interactive image registration technique with an application to stereo vision. In *Proc. of Seventh International Conference on Artificial Intelligence*, pages 674–679, 1981.
- [78] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *Computer Vision*, 56(3):221–255, 2004.
- [79] Teahyung Lee. *Algorithm-Based Efficient Approaches For Motion Estimation Systems*. PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, Georgia, U.S.A., December 2007.

-
- [80] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, October 2008.
- [81] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. In *IEEE ICIP*, volume 1, pages 493–496, September 2000.
- [82] S. C. Cain, M. M. Hayat, and E. E. Armstrong. Projection-based image registration in the presence of fixed-pattern noise. *IEEE Trans. Image Processing*, 10(12):1860–1872, December 2001.
- [83] B. Pires and P. Aguiar. Featureless global alignment of multiple images. In *IEEE ICIP*, volume 1, pages 57–60, September 2005.
- [84] X. Dai and S. Khorram. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. Geoscience and Remote Sensing*, 36(5):1566–1577, September 1998.
- [85] D. P. Roy. The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Trans. Geoscience and Remote Sensing*, 38(4):2017–2032, July 2000.
- [86] L. Bruzzone, R. Cossu, and M. Gomasasca. Adaptive estimation of the registration-noise distribution for accurate unsupervised change detection. In *IEEE IGARSS*, volume 6, pages 2584–2586, July 2001.
- [87] L. Bruzzone and R. Cossu. An unsupervised change detection technique robust to registration noise. In *IEEE IGARSS*, volume 1, pages 306–308, June 2002.
- [88] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Processing*, volume 1, pages 317–339, October 1984.
- [89] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, September 2002.
- [90] Y. Altunbasak, A. J. Patti, and R. Mersereau. Super-resolution still and video reconstruction from MPEG-coded video. *IEEE Trans. Circuits Systems Video Technology*, 12(4):217–226, April 2002.

-
- [91] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, May 2003.
- [92] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Processing*, 13(10):1327–1344, October 2004.
- [93] S. Farsiu, M. Elad, and P. Milanfar. Multiframe demosaicing and super-resolution of color images. *IEEE Trans. Image Processing*, 15(1):141–159, January 2006.
- [94] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Video-to-video dynamic superresolution for grayscale and color sequences. *Applied Signal Processing*, 2006(10):232–246, January 2006.
- [95] N. A. Ahuja and N. K. Bose. Superresolution and noise filtering using moving least squares. *IEEE Trans. Image Processing*, 15(8):2239–2248, August 2006.
- [96] A. K. Katsaggelos, R. Molina, and J. Mateos. *Super Resolution of Images and Video*. Morgan & Claypool, New York, first edition, 2007.
- [97] P. Vandewalle, L. Baboulaz, P. L. Dragotti, and M. Vetterli. Subspace-based methods for image registration and super-resolution. In *IEEE ICIP*, pages 645–648, October 2008.
- [98] M. Tanaka, Y. Yaguchi, and M. Okutomi. Robust and accurate estimation of multiple motions for whole-image super-resolution. In *IEEE ICIP*, pages 649–652, October 2008.
- [99] T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, and R. Silverman. An efficient k -means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):881–892, July 2002.
- [100] Mordecai Avriel. *Nonlinear Programming: Analysis and Methods*. Dover, U.S.A., 2003.
- [101] J. Nocedal and S. Wright. *Numerical optimization*. New York: Springer, third edition, 1999.
- [102] <http://vision.ece.ucsb.edu/registration/satellite/>.

-
- [103] M. Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup. Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158, January 2008.
- [104] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, April 2004.
- [105] Al Bovik. *Image and Video Processing*. Academic Press, Texas, second edition, 2000.
- [106] D. Barnea and H. Silverman. A class of algorithms of fast digital image registration. *IEEE Trans. Computers*, 21(2):179–186, February 1972.
- [107] <http://users.isr.ist.utl.pt/~aguiar/mosaics>.
- [108] Stan Z. Li. *Markov Random Field: Modeling in Computer Vision*. Springer, New York, U.S.A., second edition, 2000.
- [109] A.N. Tikhonov. Regularization of incorrectly posed problem. *Soviet Mathematical Dolk.*, 4(6):1624–1627, 1963.
- [110] S. Geman and D. E. McClure. Bayesian image analysis: An application to single photon emission tomography. In *Statistical Computation Section*, volume 14, pages 12–18. Amer. Statistical Assoc., 1985.
- [111] P.J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Medical Imaging*, 9:84–93, March 1990.
- [112] T. Hebert and R. Leahy. A generalized em algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Medical Imaging*, 8:194–202, June 1990.
- [113] B. Rouchouze, P. Mathieu, T. Gaidon, and M. Barlaud. Motion estimation based on markov random fields. In *IEEE ICIP*, volume 3, pages 270–274, November 1994.
- [114] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edges-preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6(2):289–311, February 1997.

-
- [115] Ricardo A. Maronna, R. Douglas Martin, and Víctor J. Yohai. *Robust Statistics: Theory and Methods*. John Wiley & Sons, Sussex, England, 2006.
- [116] P.J. Huber and R. Dutter. Numerical solution of robust regression problems. In *Proc. Symp. Computational Statistics*, pages 165–172. G. Brushmann, 1974.
- [117] D.F. Shanno and D.M. Rocke. Numerical methods for robust regression: Linear models. *SIAM J. Scientific and Statistical Computing*, 7:86–97, 1986.
- [118] H. Ekblom. A new algorithm for the Huber estimator in linear models. *BIT Numerical Mathematics*, 28(1):123–132, March 1988.
- [119] K. Madsen and H.B. Nielsen. Finite algorithms for robust linear regression. *BIT Numerical Mathematics*, 30(4):682–699, December 1990.
- [120] S.J. Wright. Using complementarity and optimization methods in statistics. In *Proc. Inter. Conf. on Complementarity Problems*, pages 165–172, June 1999.
- [121] Olvi L. Mangasarian and David R. Musicant. Robust linear and support vector regression. *IEEE Trans. Pattern Analysis And Machine Intelligence*, 22(9):950–955, September 2000.
- [122] Karl-Rudolf Koch. *Parameter estimation and hypothesis testing in linear model*. Springer-Verlag, Berlin, Germany, second edition, 1999.
- [123] John Fox. Robust regression. Technical report, Thousand Oaks/London /New Delhi: Sage, January 2002.
- [124] M. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. Image Processing*, 7(3):421–432, March 1998.

Appendix A

Real Image Pairs

In this dissertation, ten real image pairs [102] were used in the experiments as mentioned in Section 3.5. Two pairs of them are shown in Fig. 2.1(a,b) and Fig. 2.1(c,d); pair #A and pair #B, respectively. In this appendix, the other eight image pairs are depicted in Fig. A.1 through Fig. A.8; pair #C through pair #J, respectively.

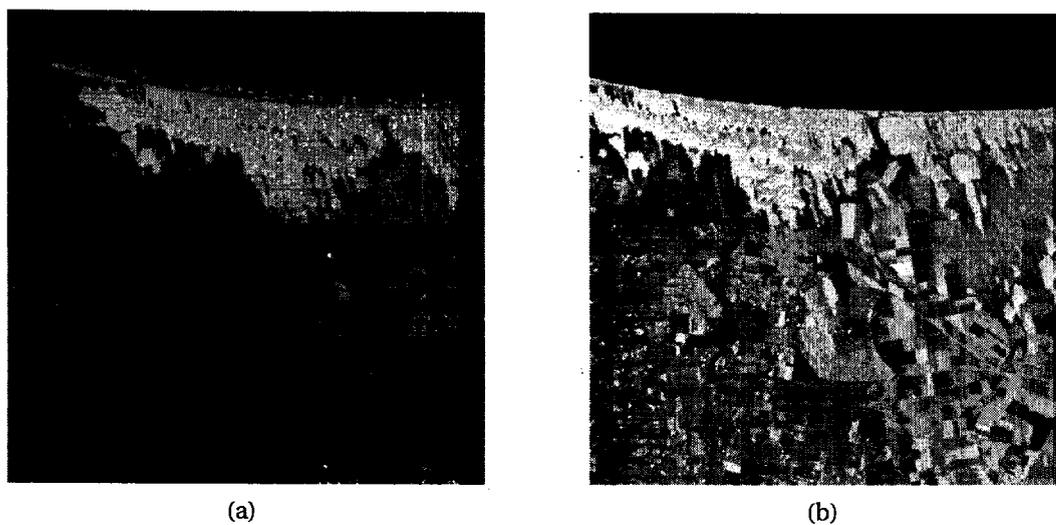


Figure A.1: Real image pair #C.

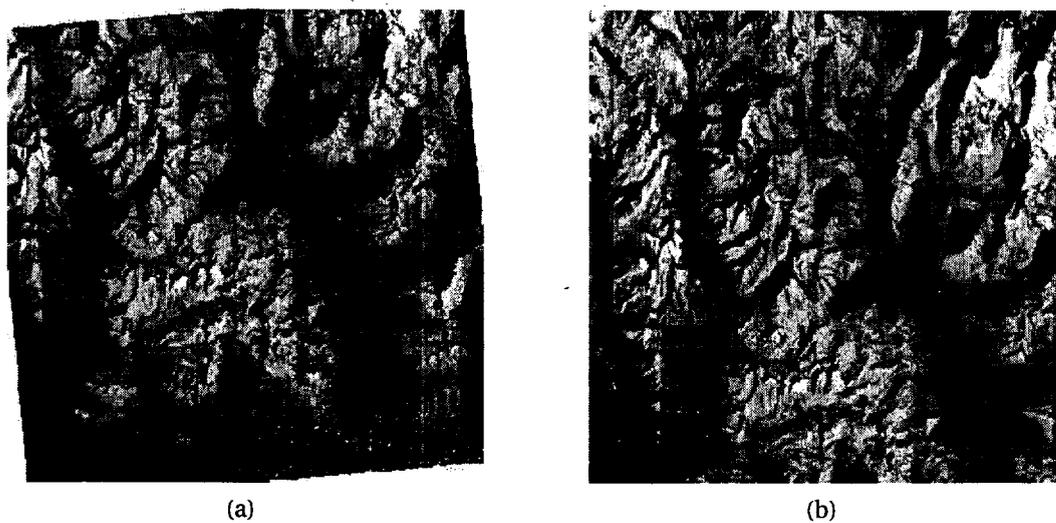
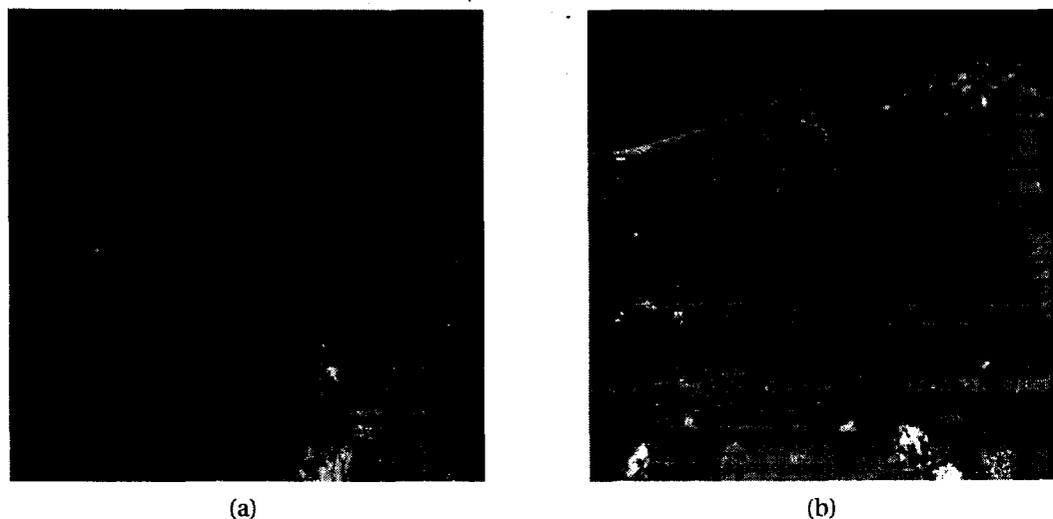


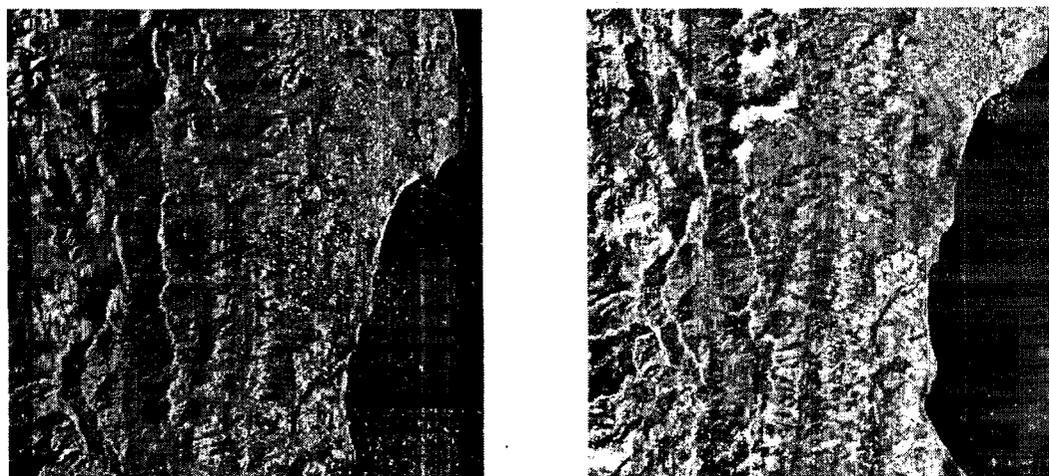
Figure A.2: Real image pair #D.



(a)

(b)

Figure A.3: Real image pair #E.



(a)

(b)

Figure A.4: Real image pair #F.

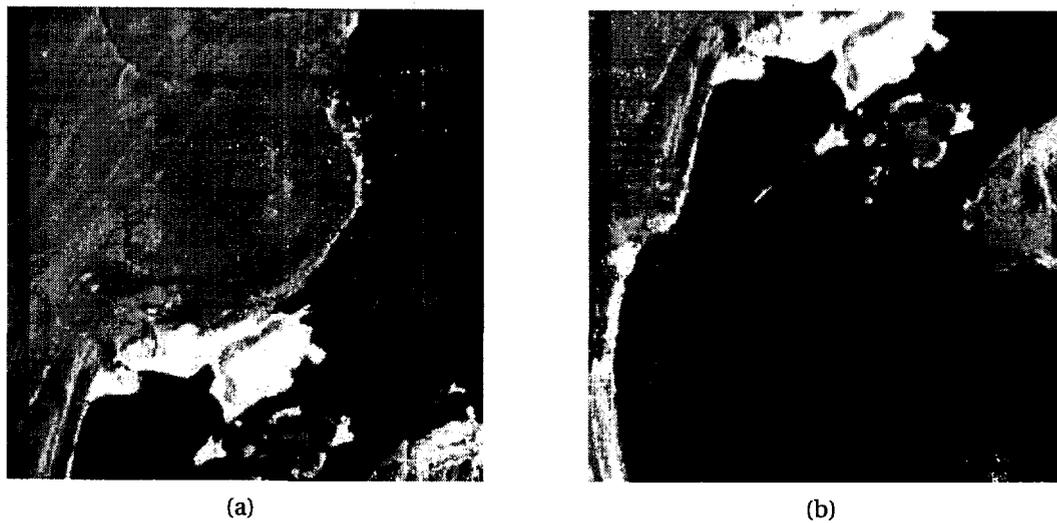


Figure A.5: Real image pair #G.

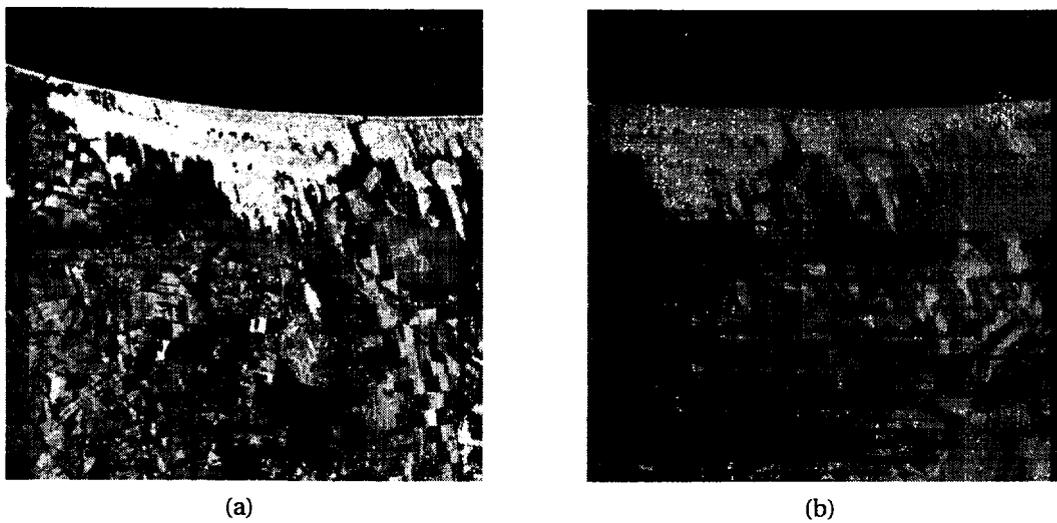


Figure A.6: Real image pair #H.

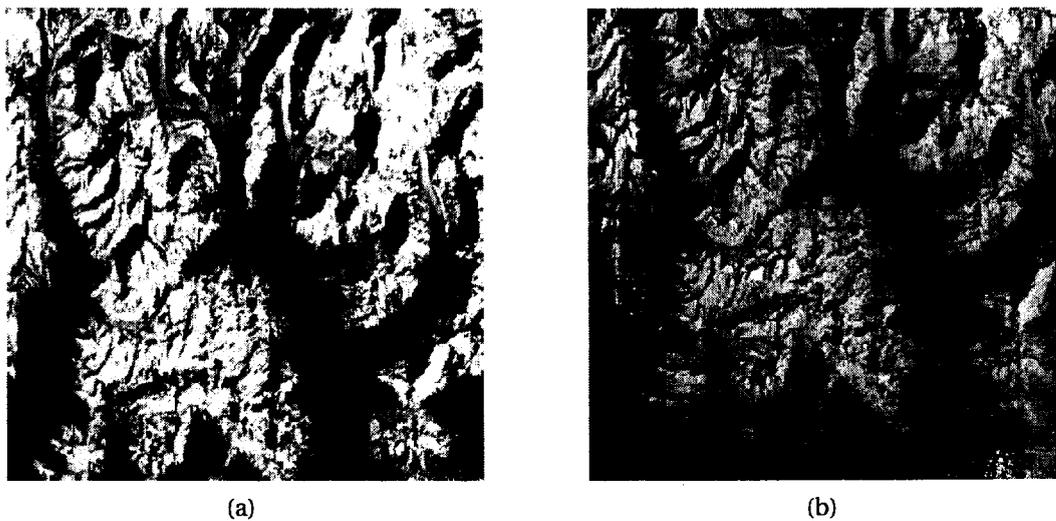


Figure A.7: Real image pair #I.

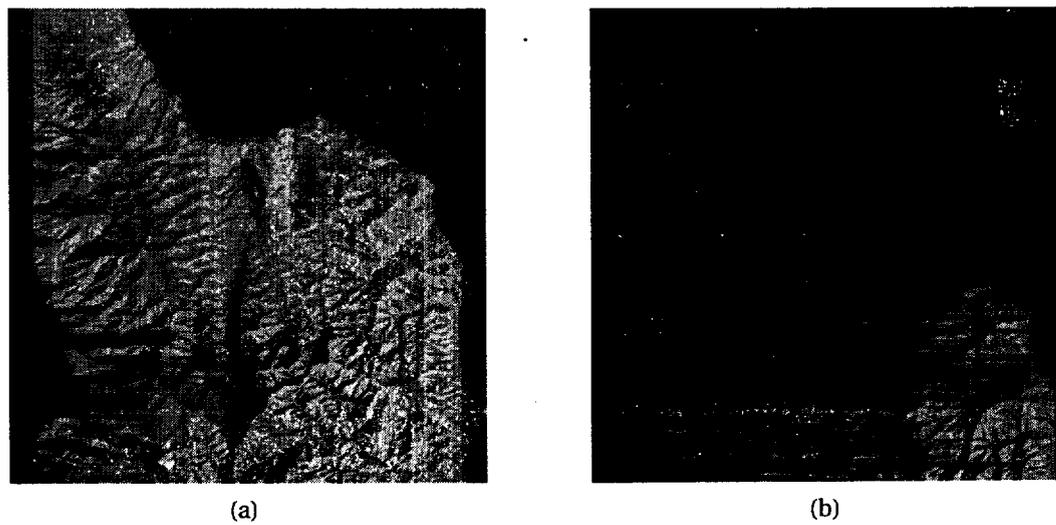


Figure A.8: Real image pair #J.

Appendix B

Segmentation Deviations Results

This appendix show Tables (B.1) through (B.6) that report the percentage increase of the AAEs of the estimated six affine parameters, respectively, for different SNLs w.r.t. those AAEs without segmentation error, *i.e.*, at SNL=0, using LS-GIM, LS-ASLIV_{6,6}, HM-ASLIV_{6,6}, RH-ASLIV_{6,6} and WH-ASLIV_{6,6} approaches for the “ $J=3$ ” data set. Depicted is that WH-ASLIV approach outperforms other approaches while SNL increases, thereby being less sensitive to segmentation perturbations than the others. The results of those six tables are depicted in Fig. 5.6, respectively. Note that the lower the percentage value of AAEs, the better the performance of the approach.

Table B.1: Percentage increase of AAEs of the estimated parameter a_1 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set. The lower the percentage value of AAEs, the better the performance of the approach.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	15.3658×10^{-4}	1.7205×10^{-4}	1.2025×10^{-4}	0.9806×10^{-4}	0.9167×10^{-4}
265	3.47%	2.88%	2.63%	2.13%	1.70%
511	5.10%	4.89%	4.57%	4.40%	3.75%
764	7.33%	7.10%	6.27%	5.78%	4.60%
907	8.59%	8.00%	7.85%	7.33%	6.50%
1031	10.27%	10.35%	10.29%	9.86%	8.95%
1146	11.77%	13.18%	13.72%	13.07%	10.73%
1353	14.59%	15.16%	17.69%	16.55%	12.61%
1708	18.30%	18.93%	20.94%	18.49%	13.93%
2016	22.53%	23.78%	23.50%	21.91%	16.48%
2187	25.84%	26.72%	26.55%	26.71%	18.51%
2354	28.15%	29.41%	28.70%	31.60%	19.73%
2627	32.84%	32.13%	32.34%	33.76%	22.00%
2795	35.90%	36.19%	35.02%	38.48%	23.41%
3124	40.36%	39.79%	37.46%	41.68%	25.38%

Table B.2: Percentage increase of AAEs of the estimated parameter a_2 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	12.0018×10^{-4}	1.5079×10^{-4}	1.0344×10^{-4}	0.8096×10^{-4}	0.7748×10^{-4}
265	3.28%	2.71%	2.54%	2.12%	1.95%
511	5.31%	4.92%	4.63%	4.49%	3.68%
764	7.25%	6.94%	6.50%	5.54%	4.36%
907	8.71%	8.14%	7.98%	7.55%	6.78%
1031	10.22%	10.27%	10.14%	8.81%	7.98%
1146	14.05%	11.77%	12.81%	11.36%	10.70%
1353	17.80%	14.70%	16.30%	14.92%	13.46%
1708	22.11%	17.23%	18.20%	17.53%	15.06%
2016	24.54%	19.60%	21.47%	20.15%	17.14%
2187	28.54%	24.20%	23.82%	22.45%	18.71%
2354	30.95%	28.67%	26.29%	25.13%	20.33%
2627	35.12%	31.87%	29.30%	28.03%	23.32%
2795	37.40%	37.41%	32.91%	31.59%	24.83%
3124	40.21%	41.14%	34.92%	34.76%	26.28%

Table B.3: Percentage increase of AAEs of the estimated parameter a_3 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set. The lower the percentage value of AAEs, the better the performance of the approach.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	14.6081×10^{-4}	1.6189×10^{-4}	1.1661×10^{-4}	0.9399×10^{-4}	0.8779×10^{-4}
265	3.51%	2.92%	2.67%	2.33%	1.87%
511	5.39%	5.13%	4.72%	4.10%	3.59%
764	7.14%	6.82%	6.41%	5.83%	4.92%
907	9.24%	8.86%	8.35%	7.69%	6.83%
1031	11.30%	10.32%	10.13%	8.93%	9.68%
1146	14.34%	13.13%	11.52%	12.17%	11.27%
1353	17.03%	15.25%	13.02%	14.30%	13.60%
1708	18.83%	17.32%	14.84%	17.69%	15.73%
2016	22.35%	20.14%	18.23%	20.82%	17.77%
2187	24.23%	22.59%	21.80%	24.75%	21.02%
2354	26.63%	24.52%	23.54%	29.19%	23.99%
2627	31.57%	26.65%	26.62%	34.27%	26.05%
2795	35.87%	31.03%	31.06%	37.69%	28.23%
3124	40.88%	34.66%	33.98%	40.14%	30.36%

Table B.4: Percentage increase of AAEs of the estimated parameter a_4 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	19.1128×10^{-4}	2.0502×10^{-4}	1.8699×10^{-4}	1.6531×10^{-4}	1.5612×10^{-4}
265	3.37%	1.90%	2.52%	2.29%	1.66%
511	5.50%	3.25%	4.82%	4.46%	3.71%
764	7.43%	5.53%	6.65%	5.63%	4.58%
907	8.90%	8.43%	7.87%	7.28%	6.42%
1031	10.05%	9.69%	10.34%	9.96%	7.81%
1146	11.65%	12.47%	13.01%	11.07%	9.31%
1353	13.84%	14.87%	16.04%	13.66%	10.89%
1708	15.93%	19.10%	18.64%	15.85%	12.01%
2016	17.94%	22.67%	21.18%	18.55%	13.53%
2187	20.97%	25.65%	24.31%	21.54%	15.21%
2354	23.77%	28.73%	25.91%	23.41%	16.57%
2627	27.13%	31.50%	30.07%	24.96%	18.66%
2795	29.68%	35.15%	33.83%	28.21%	21.37%
3124	34.04%	38.06%	35.57%	30.53%	24.14%

Table B.5: Percentage increase of AAEs of the estimated parameter a_5 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set. The lower the percentage value of AAEs, the better the performance of the approach.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	9190.77×10^{-4}	789.16×10^{-4}	722.35×10^{-4}	691.29×10^{-4}	647.22×10^{-4}
265	3.22%	2.75%	2.67%	2.21%	1.79%
511	4.94%	4.76%	4.72%	4.03%	3.63%
764	7.32%	6.88%	6.41%	5.42%	4.49%
907	8.79%	8.20%	8.00%	7.08%	6.19%
1031	11.96%	9.29%	10.34%	8.38%	7.25%
1146	15.52%	11.56%	11.49%	9.92%	9.31%
1353	17.61%	15.69%	13.23%	11.77%	11.46%
1708	21.54%	18.79%	15.77%	13.42%	13.20%
2016	24.46%	22.26%	17.54%	16.03%	15.17%
2187	29.55%	26.77%	21.27%	18.01%	17.23%
2354	31.71%	31.48%	23.63%	20.22%	18.47%
2627	36.46%	35.88%	26.64%	22.21%	20.32%
2795	40.29%	41.22%	28.60%	24.85%	22.19%
3124	42.20%	44.51%	31.67%	27.60%	23.78%

Table B.6: Percentage increase of AAEs of the estimated parameter a_6 versus different SNLs, *i.e.*, $(P_e + P_d)$ with “ $J=3$ ” data set.

SNL	LS-GIM	LS-ASLIV	HM-ASLIV	RH-ASLIV	WH-ASLIV
0	2688.72×10^{-4}	795.70×10^{-4}	716.11×10^{-4}	685.77×10^{-4}	644.32×10^{-4}
265	3.31%	1.90%	2.55%	2.07%	1.68%
511	4.87%	3.25%	4.26%	3.89%	3.40%
764	7.03%	5.53%	5.89%	5.31%	4.73%
907	8.93%	8.43%	7.85%	7.44%	6.36%
1031	11.10%	10.09%	9.35%	9.86%	7.92%
1146	14.15%	11.43%	11.17%	11.13%	9.47%
1353	16.27%	14.98%	13.09%	12.86%	10.58%
1708	18.31%	18.40%	16.03%	14.52%	12.78%
2016	21.84%	22.98%	18.40%	16.59%	14.15%
2187	24.80%	26.96%	20.70%	19.11%	15.90%
2354	28.14%	29.45%	22.80%	21.62%	17.16%
2627	30.18%	33.95%	25.96%	23.89%	19.31%
2795	34.85%	36.46%	27.54%	25.16%	21.47%
3124	37.68%	41.52%	30.51%	27.72%	22.83%